

Brand Pitt: A Corpus to Explore the Art of Naming

Gözde Özbal, Carlo Strapparava, Marco Guerini

FBK-Irst, Trento-Italy

Abstract

The name of a company or a brand is the key element to a successful business. A good name is able to state the area of competition and communicate the promise given to customers by evoking semantic associations. Although various resources provide distinct tips for inventing creative names, little research was carried out to investigate the linguistic aspects behind the naming mechanism. Besides, there might be latent methods that copywriters unconsciously use. In this paper, we describe the annotation task that we have conducted on a dataset of creative names collected from various resources to create a gold standard for linguistic creativity in naming. Based on the annotations, we compile common and latent methods of naming and explore the correlations among linguistic devices, provoked effects and business domains. This resource represents a starting point for a corpus based approach to explore the art of naming.

Keywords: linguistic creativity, naming, branding

1. Introduction

A catchy, memorable and creative name is one of the most important keys to a successful business since the name provides the first image and defines the identity of the service to be promoted. A good name is able to state the area of competition and communicate the promise given to customers by evoking semantic associations. However, coming up with such a name is not easy and it is quite time consuming due to the restriction that the task imposes to fulfill all these objectives at once. Besides, since many new products and companies emerge every year and most of the obvious words have been already claimed, the naming style is continuously changing and creativity is a fundamental dimension to be considered in the naming process.

The creation of a name is both an art and a science (Keller, 2003). Effective names do not come out of the blue and for each of them precise methodologies are used. Although it might not be easy to perceive all the effort required to come up with the final output, both a training phase and a long process -consisting of many iterations- are necessary for creating a good name. Despite the fact that various resources provide distinct tips for inventing creative, attractive and memorable names, no attempt has been made to combine all means of creativity that can be used during the naming process. Furthermore, we suspect that in addition to the devices stated by copywriters, other latent methods might be unconsciously employed. For these reasons, it is essential to discover and accumulate all crucial features of creativity, since we plan to automatize the naming process eventually.

In this paper, we describe an annotation task that we have conducted on a dataset of 1000 brand and company names collected from various resources to create a gold standard for linguistic creativity in naming. This task has required a total of three linguists to determine both the creativity devices used in each name and the effects that these names provoke in order to find out how the names are actually

perceived.

Based on the resulting annotations, we compile common and latent methods of naming and discover the correlations between various features. We plan to exploit this compilation to bootstrap the automatization of the naming process in a more systematic way than previous attempts. Besides, this dataset can be utilized for learning a model to identify creative words. Such a model would be extremely useful for automatically getting feedback for words invented by non-professionals and even naming agents.

The rest of the paper is structured as follows. We review the state-of-the-art relevant to the naming task in Section 2. We explain the annotation procedure that we have followed in Section 3. We describe the annotated dataset in Section 4. We present preliminary feature correlation analysis in Section 5. Finally, we draw conclusions and outline ideas for possible future work in Section 6.

2. Background

Currently, many naming agencies and branding firms (e.g. www.designbridge.com, eatmywords.com, ahundredmonkeys.com) provide professional service to aid with the naming of new products, domains, companies and brands. Such services generally require customers to provide brief information about the business to be named, fill in questionnaires to learn about their markets, competitors and expectations. In the end, they present a list of names to be chosen from. Although the resulting names can be successful and satisfactory, these services are expensive and the processing time is rather long.

As an alternative, automatic name generators can be used as a source of inspiration. For instance, www.business-name-generators.com randomly combines abbreviations, syllables and generic short words from different domains to obtain creative combinations. The domain generator provided by www.namestation.com ran-

domly generates name ideas and available domains based on alliterations, compound words and custom wordlists. It allows users to determine the prefix and suffix of the name to be generated. The brand name generator on www.netsubstance.com takes as inputs keywords and lets users configure the percentage of the shifting of keyword letters. Lastly, www.naming.net is based on name combinations among common words, Greek and Latin prefixes, suffixes and roots, beginning and ending word parts and rhymes. A shortcoming of these systems is that they all use random generation which can result in many bad suggestions and users have to be patient to find a satisfactory name. Besides, they are only based on straightforward combinations of words and no mechanism exists to also include semantic reasoning.

Little research has been carried out to investigate the linguistic aspects of the naming mechanism. B. V. Bergh (1987) built a four-fold linguistic topology consisting of phonetic, orthographic, morphological and semantic categories to evaluate the frequency of linguistic devices in brand names. Bao et al. (2008) investigated the effects of relevance, connotation, and pronunciation of brand names on preferences of consumers. Klink (2003) focused on the relationship between structural characteristics of brand names and visual characteristics of brand marks. He also explored ways of associating brand names and brand marks to communicate consistent brand meaning. Klink (2000) based its research on the area of sound symbolism (i.e. “the direct linkage between sound and meaning” (Leanne Hinton, 2006)) by investigating whether the sound of a brand name conveys an inherent meaning and the findings showed that both vowels and consonants of brand names communicate information related to products when no marketing communications are available. Kohli et al. (2005) analyzed consumer evaluations of meaningful and non-meaningful brand names and the results suggested that non-meaningful brand names are evaluated less favorably than meaningful ones even after repeated exposure. Lastly, Hernández (2011) presented a recent study that focused on the semantics of branding. Based on the analysis of several international brand names, it was shown that cognitive operations such as domain reduction/expansion, mitigation, and strengthening might be used unconsciously while creating a new brand name.

3. Annotation

To build our dataset, we conducted four phases of annotation tasks using a total of three linguists to annotate various information for 1000 brand and company names. Although our first intention was to use the crowdsourcing service of Mechanical Turk¹, we excluded this option mainly due to the difficulty of the task since the expert knowledge required was very high. In addition, by avoiding random annotations and using expert linguists trained in the task, we have obtained a much more reliable resource.

In the first phase of the annotation task, each annotator was required to fill in the business domain description together with the etymology for one third of the 1000 names in the dataset. As an example, for the brand name *Kotex*, the business domain was stated to be “sanitary towels” and the etymology was determined simply as “cotton texture” while for the name *Samsung*, the business domain and etymology were annotated as “electronics” and “three stars in Korean” respectively. During this phase, the annotators were free to use any kind of resource. The business domain and etymology fields provided very useful information which would help the annotators to realize the creativity devices used for the names more easily and correctly in the second phase. At the end of the first phase, all three annotators checked the two fields that they had annotated for each name and made the required changes based on agreement whenever necessary. While the annotators were initially allowed to input free text for the business domain field, later on two of the annotators manually processed all the entries in this field to uniform the data as explained in detail in Section 4.

The second annotation phase required the determination of the creativity devices present in each name - by considering the business domain and etymology information - together with the effects that these names provoke. The list of these creativity devices and effects are explained in detail in Section 4. The annotators were allowed to choose more than one creativity device and effect for each name. In this phase, one annotator was responsible for the annotation of the whole data set while the other two were asked to annotate half of it. For the creativity devices, we provided a list of attributes that we collected from various resources. We explained these attributes with examples in the annotation guidelines. Additionally, we provided another field with the name *Other* so that annotators could state any other device that was not explicitly listed. The information in this field has great value for our goal, since it eventually helped us to discover the latent creativity devices.

After the second phase, the annotators took role in the *reconciliation phase* for the annotation of the creativity devices. In this phase, for the cases with disagreement they were able to see the answers of each other and to make changes in their own answers whenever they agreed with the other annotator. If they still thought that their own answer was correct, they added a note justifying their answer. During the reconciliation phase, only the annotations of the creativity devices were taken into consideration since the creativity effects are subjective, names can be perceived differently and we do not expect to have one correct answer set for these fields. The inter-annotator agreement of the annotators for the creativity devices measured using Cohen’s Kappa after the reconciliation phase can be found in Table 1. In this table, the categories are listed in the first column while the range of the Kappa values are stated in the second column. As can be observed from the table, the agreement among the annotators is very high and the creativity device with least agreement is *personification*.

The last phase of our annotation procedure was the *final ad-*

¹<https://www.mturk.com/mturk>

Creativity Device	Kappa
Oxymoron, Plosives, Acronym, Palindrome, Fictional, Historical, Location, Latin	1
Founder name, Onomatopia, Clipping and Blending, Assonance, Consonance, Invented, Juxtaposition	0,99 - 0,98
Abbreviation, Strengthening, Punning, Plant/Animal names, Person/character, Mythical, Functional, Descriptive, Unusual/incorrect spelling	0,97 - 0,96
Reduplication, Semantic, Affixation, Mitigation, Rhyme, Alliteration, Metaphor	0,95 - 0,93
Personification	0,89

Table 1: Inter-annotator agreement for various creativity devices using Cohen’s Kappa

judication where a third annotator made the final decision for the annotations of the creativity devices with disagreement. Obviously, the third annotator for each half of the dataset was the one who did not take role in the annotation of that half in the previous phases. By checking the notes of the other two annotators justifying their own decisions, the third annotator agreed with one of them and made the final decision so that all the conflicts in the dataset were resolved.

4. Dataset

For the annotation task, we collected 1000 brand and company names from several different domains. These names were retrieved from a book dedicated to brand naming strategies (Botton and Cegarra, 1990) and various web resources related to creative names such as `adslogans.co.uk` and `brandsandtags.com`.

Our list contains names which were invented according to various creativity methods. While the creativity in some of these names is independent of the context and the names themselves are sufficient to realize the methods used (e.g. alliteration in *Peak Performance*, modification of one letter in *Vimeo*), for some of them the context information such as the description of the product or the area of the company is also necessary to fully understand the methods used. For instance, *Thanks a Latte* is a coffee bar name where the phonetic similarity between “lot” and “latte” (a coffee type meaning “milk” in Italian) is exploited. In the name *Caterpillar*, which is an earth-moving equipment company, provocation is used. Therefore, we need extra information regarding the domain description in addition to the names while determining the creativity devices.

In the remainder of this section, we will describe in detail the creativity devices and effects that we have considered for the annotation procedure. We will also list the latent devices that we did not initially include in the annotation but we discovered by analyzing the *Other* field that was used whenever the annotators found out a different technique from the ones already provided.

4.1. Creativity devices in naming

We collected a total of 31 attributes used in the naming process from various resources including academic papers, naming agents, branding and advertisement experts. To facilitate the task for the annotators, we subsumed the most similar attributes when required. Adopting the four-fold linguistic topology suggested by (B. V. Bergh, 1987), we mapped most of these attributes into *phonetic*, *orthographic*, *morphological* and *semantic* categories. We placed the remaining devices into the three new categories that we created, namely *proper names*, *languages* and *plant and animal names*.

Phonetic. The phonetic category includes the devices *rhyme* (i.e. identical pronunciation of two words from the vowel in the stressed syllable to the end - e.g. *Etch-a-sketch*), *alliteration* (i.e. repetition of initial sounds in neighboring words - e.g. *Peak Performance*), *assonance* (i.e. repetition of vowel sounds - e.g. *Volvo*), *consonance* (i.e. repetition of consonant sounds - e.g. *Cookeen*), *reduplication* (i.e. repeating the root or stem of a word or part of it exactly or with a slight change - e.g. *Juicy Juice*), *onomatopia* (i.e. using a word that imitates or suggests the source of the sound that it describes - e.g. *Bark*) and *plosives* (i.e. using a “p”, “t”, “k”, “b”, “d” or “g” sound that causes to build up air pressure in the mouth and forcefully pronounce the sound - e.g. *Coca cola*).

Orthographic. The *orthographic* category consists of *unusual or incorrect spelling* (e.g. M-azing, Krispy Kreme), *abbreviation* (e.g. Dr. Eutker), *acronym* (e.g. *BMW*) and *palindrome* (i.e. words, phrases, numbers that can be read the same way in either direction e.g. *Honda “Civic”*).

Morphology. The third category is the *morphology* which contains *affixation* (i.e. forming different words by adding morphemes at the beginning, middle or end of words - e.g. *Nutella*), *compounding* (e.g. *Facebook*), *clipping* (i.e. reduction of a word to one of its parts (Marchand, 1969) - e.g. *Intelsat* from International Telecommunications Satellite Organization) and *blending* (i.e. forming a word by blending sounds from two or more distinct words and combining their meanings - e.g. *Wikipedia* by blending “Wiki” and “encyclopedia”). It should be noted that we kept the annotations for clipping and blending together since there is a very subtle difference between clipping compounds

and blending (i.e. the forms retaining compound stress are clipped compounds while those taking simple word stress are blends according to (Bauer, 1983)) and from a computational point of view this difference is not very interesting.

Semantics. The fourth category is the *semantics* which includes the devices *functional descriptive* (i.e. purely describing what a company or product does - e.g. *Turkish airlines*), *personification* (i.e. any attribution of human characteristics to animals, non-living things, phenomena, material states, objects or abstract concepts - e.g. *Mr. Muscle* for a detergent), *metaphor* (i.e. expressing an idea through the image of another object - e.g. the first popular web browser called *Mosaic* by likening the web and web page to a familiar kind of picture made out of little pieces), *metonymy* (i.e. calling a thing or concept not by its own name, but by the name of something intimately associated with that thing or concept - e.g. *Xerox* for photocopy) and *synecdoche* (i.e. using a part for the whole or the whole for a part, the special for the general or the general for the special - e.g. a large office supply chain store called *Staples*), *semantic appropriateness* (i.e. having a kind of semantic relationship with the product or organization of the company which cannot be categorized as the other devices in the semantic category - e.g. a software company called *Six Apart* due to the fact that company co-founders were born six days apart), *punning* (i.e. using a word in different senses or words with sound similarity to achieve specific effects such as humor - e.g. *Thai Me Up* for a Thai restaurant), *oxymoron* (i.e. combining contradictory terms - e.g. *Krispy Kreme* doughnuts), *mitigation* (i.e. lessening intensity or severity and causing the product to be considered as something desirable because of its harmless and easily controllable nature (Hernández, 2011) e.g. *Chevrolet*, *Smarties*) and *strengthening* (i.e. drawing attention to the dimensions and strengths of the product by also turning it into something desirable due to its attractive and/or reliable nature (Hernández, 2011) - e.g. *Burger King*, *Royal Mail*). It is worth mentioning that within the semantic category, we have kept the rhetorical devices metaphor, metonymy and synecdoche together, since they were highly associated to each other and the annotators found it difficult to differentiate them from each other. In addition, from a computational point of view, it would be quite difficult and complex to model these devices distinctively.

Proper names. The fifth category is the *proper names* which include the subcategories *founder names* (e.g. *Dolce Gabbana*), *fictional* (e.g. anti-theft protection for bikes called *Kryptonite*, which is a fictional material from the Superman saga), *mythical* (e.g. high-fashion luxury-goods manufacturer with the name *Hermes*, which is a messenger of the gods in Greek mythology and a guide to the underworld), *historical* (e.g. the search engine called *Magellan*), *person/character* (e.g. *du Maurier* cigarettes named after Sir Gerald du Maurier, a British actor and manager) and *location* (e.g. “Texas” Instruments).

Languages. We also kept track of the languages that the names belong to under the sixth category called *languages*.

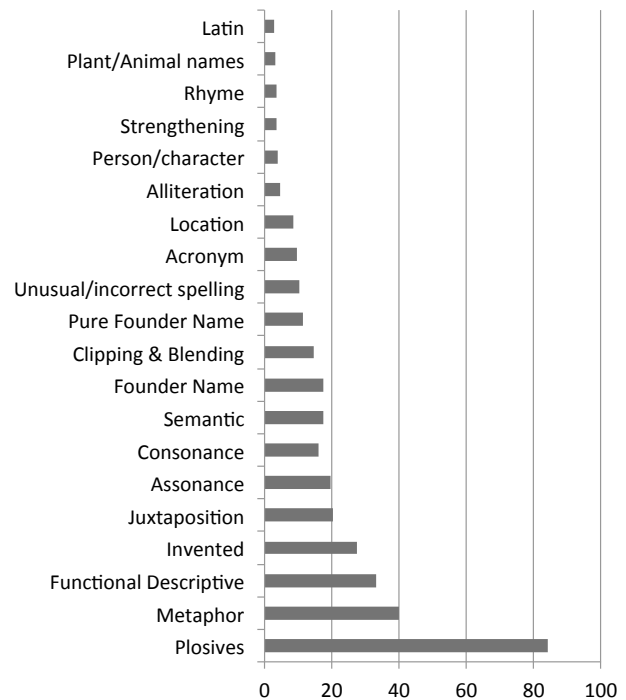


Figure 1: Corpus coverage of the main creativity devices

Before the annotation procedure, this category included only the field *Latin* (e.g. Men and women cloth store called *Aquascutum* which is a combination of two Latin words “Aqua” meaning “water” and “Scutum” meaning “shield”) and *Other* for any other language so that we could discover the names in other languages such as *French* (e.g. *Lacoste*, *Cacharel*), *Italian* (e.g. *Barilla*), *Greek* (e.g. *Eidos*), *Hebrew* (e.g. *Ariel*), *Romany* (e.g. *Tazo*), *Chinese* (e.g. *Typhoo*), *Slavic* (e.g. *Lada*) and *Japanese* (e.g. *Nissan*).

Plant and animal names. We had this category for the names using the biological or commonly used names of plants and animals. This category includes names such as *Palmolive* and *Camel*.

In Figure 1, we provide the percentage of the corpus coverage of the most frequent creativity devices (with a frequency higher than 3%). As can be noticed from this figure, a very high portion of the names (84.1%) include plosives. Interestingly, metaphor is the second most frequently referred device in the dataset (39.8%).

As we expected, the *Other* field, which allowed the annotators to state any additional creativity device they observed for the names, was indeed useful to discover new devices that can be used in the naming process. The list of the latent devices that we discovered by analyzing this field are as follows:

- Using positively connotated words (e.g. *Cheer*, *Cat Fancy* and *Lucky*)
- Using negatively connotated or provocative words (e.g. *Ugg* as a modification of “ugly”, *Obey Clothing* for a clothing store, *Mania*)

- *Latinizing* a non-Latin word by concatenating a Latin suffix to it (e.g. *Samsonite*)
- Word play obtained by representing the letter or string repetition with a number (e.g. *3M* for Minnesota Mining and Manufacturing Company and *3Com* representing computer, communication, and compatibility)
- Using negation (e.g. *Not Your Daughter’s Jeans, No Rinse, I Can’t Believe It’s Not Butter*)
- Using a sentence form (e.g. *Obey Clothing, Volvo* which means “I roll” in Latin, *Kiss My Face*)
- Targeting/addressing the potential customer profile (e.g. *Rich and Skinny* for a jeans brand, *Elle* which means “she” in French for a women’s fashion magazine, *Forever 21* for clothing retailers)
- Tautonymy: Creating a brand name consisting of two identical parts (e.g. *XOXO, Yo-Yo*)
- Paedonymy: Naming a product/company after one’s child (e.g. Fast food chain restaurant *Wendy’s* named after the founder’s fourth child Melinda Lou “Wendy” Thomas.)
- Inventing names using *sound patterns* (A permissible combination of distinctive sounds or phonemes in a given language) of another language (e.g. *Hägen-Dazs* which is a non-sensical name designed specifically to make Americans think that it is Scandinavian)
- Colloquialism (e.g. *Cracker Jack* meaning “of excellent quality”)
- Spelling an already existing name backwards (e.g. *Harpo Productions* founded by “Oprah” Winfrey, *Xobni* for “inbox”)
- Using the translation of the founder’s name (e.g. *Bridgestone* - the surname of the founder Ishibashi means “stone bridge”, or “bridge of stone”)
- Using an exclamation mark (e.g. *JOOP!, Yahoo!, Yelp!*)
- Using a question mark (e.g. *Guess?*)
- Using imperative form (e.g. *Envy, Obey, Cheez-It*)
- Using a word having a substring related to the topic (e.g. *ConfIDENT* for a tooth whitening gel, *BeneFit* for a skincare and cosmetics company)
- Using the enunciation of an acronym (e.g. *Esso* for S.O. in Standard Oil)
- Adding vowels to an acronym to make the pronunciation easier (e.g. *Becel* originates from the three letter acronym BCL (Blood Cholesterol Lowering), *TIGI* from founder names: Toni&Guy)
- Spelling the pronunciation (e.g. *Reddi-Wip* for ready + whip)
- Shortening a substring with a letter based on the phonetic similarity (e.g. *M-azing*)
- Using a placeholder name (e.g. *Whatchamacallit*)
- Anagram (e.g. *Ferodo* using an anagram of the name of its founder Herbert Froode, *Vimeo* as an anagram of “movie”)
- Interjection (e.g. *Yahoo!*)
- Exaggeration (e.g. *Forever 21, Make Up For Ever, Ray-Ban* for sunglasses)
- Vulgarism (e.g. *FCUK*)
- Antithesis (e.g. An aftermarket vehicle tracking system called *Lojack* as an antithesis of “hijack”)

Domain	Corpus Coverage
food	22.90%
computing	17.50%
clothing	10.10%
personal_care	9.30%
beverages	8.80%
transport	6.80%
media	4.90%
retail	3.80%
household	3.60%
finance	1.80%
health	1.40%
others	9.10%
total	100%

Table 2: Brand domain coverage

Sub Domain	Corpus Coverage
beverages; beer_and_liquor	3.70%
beverages; carbonated_soft_drinks	1.10%
beverages; coffee	0.40%
beverages; juice	0.90%
beverages; tea	1.10%

Table 3: Brand sub-domain coverage example

4.2. Creativity effects

In addition to the creativity devices referring to the recipe for naming, we also investigated the effects that the names in our dataset provoked or how the names were actually perceived by the audience. Accordingly, we added four main questions to explore the evocative power of those names:

- **Qualities:** We asked the annotators if the names conveyed information about their size, sharpness, lightness, intensity, thickness, softness, speed, temperature, taste, gender, friendliness, strength, weight and beauty. These are commonly used attributes to study the perceived qualities of brand names (Leanne Hinton, 2006).
- **Emotion:** Based on the statement that words arousing stronger emotions achieve higher levels of awareness than less emotionally arousing words (Bock, 1986), the annotators were required to state whether the names triggered any positive or negative emotion.
- **Humor:** Humor is a popular strategy used in advertising (Stock and Strapparava, 2003) and our list includes devices which might produce humorous effects such as punning, oxymorons or metaphors. Therefore, we investigated whether the names were indeed found to be humorous.
- **Color:** Based on the previous research about *synaesthesia* (i.e. the automatical triggering of a perception in a second modality by the stimulation of one sensory modality without any direct stimulation to this

Creativity Devices	Effects					Business domains				
	Sux	Hum	Pos	Neg	Color	Food	Bev	Cloth	Care	Comp
Semantic appropriateness	0.93	0.28	0.71	0.13	0.41	0.21	0.09	0.10	0.09	0.22
Pure founder name	0.52	0	0.43	0.12	0.36	0.22	0.15	0.13	0.08	0.04
Functional Descriptive	0.75	0.20	0.55	0.06	0.45	0.26	0.06	0.05	0.06	0.22
Personification	0.85	0.44	0.78	0.04	0.41	0.37	0.11	0.04	0.00	0.11
Metaphor	0.93	0.31	0.69	0.16	0.63	0.24	0.09	0.13	0.12	0.19
Punning	0.96	0.88	0.80	0.12	0.64	0.36	0.20	0.00	0.08	0.16
Rhyme	0.86	0.43	0.77	0.20	0.63	0.43	0.06	0.26	0.03	0.00
Reduplication	1.00	0.62	0.90	0.09	0.71	0.48	0.10	0.14	0.05	0.00
Onomatopia	1.00	0.52	0.76	0.14	0.48	0.52	0.05	0.10	0.05	0.14
Unusual/incorrect spelling	0.95	0.24	0.65	0.04	0.42	0.28	0.08	0.08	0.13	0.19
Acronym	0.68	0.06	0.28	0.07	0.24	0.11	0.02	0.05	0.01	0.37
Clipping & Blending	0.93	0.11	0.55	0.07	0.37	0.14	0.06	0.04	0.09	0.34

Table 4: Feature correlation analysis among creativity devices, effects and business domains

second modality (John E. Harrison, 1997)), we explored whether implicitly associating a name to a specific color is a method used in naming. Accordingly, we asked if the names in the dataset brought a color to their minds.

- **Success:** Finally, we asked the annotators whether they found the name a successful choice for the related business domain or not. To be able to rule out the possibility that the judgment was based on any previous exposure (e.g. tv commercials), we also asked if they already knew that name.

4.3. Domain Labeling

To facilitate the analysis of brand names, a last phase of the annotation task required the annotators to map the business domain description into a small set of domain labels (e.g. `food`, `clothing`). These labels are consistent with the ones used in *SloGun*, a corpus containing corporate and political slogans along with the brand which they refer to (Guerini et al., 2008). For those names that do not have a corresponding label in *SloGun* we took the definition from Wikipedia. This can allow us to make cross comparison of brand names and slogans according to the domain of interest in the future.

Business domains were annotated according to broad commercial categories and they were split into sub-domains - if possible - to better specify their commercial or product sector. In the dataset, semicolons divide the domain from the sub-domain (e.g. `personal_care`; `cosmetics`). For those cases in which a product or service covers more than on category, the domain labels are divided by commas (e.g. `health`, `education`). Such way of annotating business domains, which does not refer to a strict ontology allows us to find correlations between creativity devices and product domains at multiple levels. In Table 2 we provide the top-domain labels used in the corpus along with their coverage. In Table 3 we provide the subtree for the `beverages` node.

5. Data Analysis

In this section, we show an example analysis of the correlations emerging from the dataset in order to give a flavor of the interesting phenomena that can be observed in this resource.

In Table 4, we demonstrate the correlations of a subset of creativity devices listed in the first column with the creativity effects under the block of columns labeled as *Effects* and a set of business domains under the block of columns labeled as *Business domains*. The creativity effects considered in this analysis are success, humor, positive and negative emotions and color which are represented as *Sux*, *Hum*, *Pos*, *Neg* and *Color* respectively under the *Effects* column block. For the business domains, we have only selected five of them with the highest corpus coverage, namely food, beverages, clothing, personal_care and computing represented as *Food*, *Bev*, *Cloth*, *Care*, *Comp*. Each entry under the *Effects* column block is simply the number of names which were created according to the corresponding creativity device and which triggered the corresponding effect on at least one annotator divided by the total number of names with that specific creativity device. Similarly, the entries under the *Business domains* column block are obtained by normalizing the number of names which were created according to the corresponding creativity device and which belong to the corresponding business domain by the same number as in the previous case.

As can be observed from the table, names consisting of pure founder names have the least success rate, while the ones produced with reduplication and onomatopia are found to be extremely successful. As expected, names based on pure founder names or acronyms are not effective choices for creating humor whereas punning is the best device to make your target audience smile. Reduplication triggers positive emotions with the highest percentage and unsurprisingly few names in the dataset convey negative emotions regardless of the creativity device used.

As for the analysis regarding the business domains, onomo-

topia and punning are the most frequently used creativity devices among the names related to food and beverage domains respectively. For clothing, rhyming has been most highly exploited while no names with puns could be found in this domain. Although there was not a very dominant method used in the names of personal care products, unusual and incorrect spelling is the most frequently referred one and no examples with personification exist in this domain. Lastly, for the computing domain, acronyms together with clipping and blending are the two most popular devices whereas rhyming and reduplication methods have not been exploited at all.

This kind of analysis can be a good starting point for someone who is willing to name his/her company or product to realize which devices would be most appropriate to use for a specific domain or to trigger a certain effect on the target customers. Considering the number of creativity features and devices that we provide in our dataset, there are still many other possible analysis to be conducted as future work.

6. Conclusion

In this paper, we have focused on the task of naming and its importance for building a successful business. We have summarized the current approaches for obtaining new names and the state of the art in this area. Afterwards, we have described the annotation task that we conducted. In this task, three linguists first determined the business domain and etymology information of 1000 brand and company names. Then, they annotated the creativity devices used for each name and determined a list of effects that these names provoke.

After analyzing the resource that we have obtained, we have evaluated the frequency of linguistic devices in our dataset. In addition, we have produced a gold standard for naming by collecting common creativity devices from various resources and by discovering latent features. Furthermore, we have presented a preliminary analysis regarding the correlations of a subset of creativity devices with the provoked effects and various business domains.

Our dataset can serve as a good starting point for a corpus based approach to explore the art of naming. As a next step, we first plan to conduct a detailed correlation analysis among all the features that we have introduced. Then, we plan to use the list of creativity devices that we have obtained to support the automatization of naming in a systematic way by combining lexical, morphological and semantic knowledge.

7. Acknowledgments

The authors were partially supported by a Google Research Award.

8. References

- L. Oliver B. V. Bergh, K. Adler. 1987. Linguistic distinction among top brand names. *Journal of Advertising Research*.
- Yeqing Bao, Alan T Shao, and Drew Rivers. 2008. Creating new brand names: Effects of relevance, connotation, and pronunciation. *Journal of Advertising Research*, 48(1):148.
- Laurie Bauer, editor. 1983. *English Word-Formation*. Cambridge University Press, Oxford.
- Michael Bock. 1986. The influence of emotional meaning on the recall of words processed for form or self-reference. *Psychological Research*, 48(2):107–112.
- Marcel Botton and Jean-Jack Cegarra, editors. 1990. *Le nom de marque*. Paris McGraw Hill.
- M. Guerini, C. Strapparava, and O. Stock. 2008. Resources for persuasion. In *Proceedings of LREC*.
- Lorena Pérez Hernández. 2011. Cognitive tools for successful branding. *Applied Linguistics*, 32:369–388.
- Simon Baron-Cohen John E. Harrison. 1997. Synaesthesia: An introduction. In *Synaesthesia: Classic and contemporary readings*, pages 3–16. Blackwell Publishers, Inc.
- Kevin Lane Keller. 2003. *Strategic brand management: building, measuring and managing brand equity*. New Jersey: Prentice Hall.
- Richard R. Klink. 2000. Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, 11(1):5–20.
- Richard R Klink. 2003. Creating meaningful brands : The relationship between brand name and brand mark. *Marketing Letters*, 14(3):143–157.
- C Kohli, K Harich, and Lance Leuthesser. 2005. Creating brand identity: a study of evaluation of new brand names. *Journal of Business Research*, 58(11):1506–1515.
- John J. Ohala Leanne Hinton, Johanna Nichols. 2006. *Sound Symbolism*. Cambridge University Press.
- Hans Marchand, editor. 1969. *The Categories and Types of Present-Day English Word-formation*. Mnchen: C.H.Beck'sche Verlagsbuchhandlung.
- Oliviero Stock and Carlo Strapparava. 2003. Getting serious about the development of computational humor. In *Proceedings of the 18th international joint conference on Artificial intelligence*, pages 59–64, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.