Lexical Combinatorial Creativity with "Gastronaut"

Cristina Butnariu [1] and **Tony Veale** [1]

Abstract. Combinatorial creativity involves the creation of new ideas (and, if realized, artefacts) by exploring atypical combinations of familiar ideas. In this paper we explore the relationship between word-level creativity and artefact-level creativity in a computational system dubbed "Gastronaut" that explores the food domain. Cookery is both a paradigm example, and a metaphor, for combinatorial creativity, since creativity can be viewed as the insightful combination of ingredients, abstract or otherwise, to generate a novel artefact. Our model uses lexical descriptions for generated ingredient combinations and a corpus approach to assess H-Creativity or P-Creativity values to reflect the degree of novelty involved in each combination. We argue that P-Creativity and H-Creativity must be seen as a graded rather than binary phenomena.

1 INTRODUCTION

In this paper we propose a computational model which captures the process of combinatorial creativity in the culinary domain. A chef combines ingredients to create novel dishes that challenge expectations but which ultimately please the taste-buds [3,5]. Not every combination of basic ideas can be meaningful, of course; to be judged as creative, the combination must be faithful to the complementary notions of Novelty and Quality [1,2]. A dish is novel if it has never been proposed before, but novelty alone does not prevent such a dish from being either unsavory or downright toxic. Quality, then, is a metric that ensures that the creative output serves the intended (or related) purpose of its creator.

Cookery is a fertile domain to explore, as the products of culinary creativity have both a lexical and a physical instantiation. The label "chocolate pizza with almonds" describes a relatively novel combination of real ingredients, but it is itself a relatively novel combination of words. Though one might expect the label to follow the product, we can easily turn this expectation on its head. By generating novel labels for dishes, the semantics of language ensure that these labels will in turn describe novel artefacts. That is, to the extent that language describes reality, we can model the creative production of novel artefacts as the creative production of novel language strings, or more specifically, noun phrases. This allows us to employ the traditional tools of language processing - lexica and grammars - in the service of creative computation.

Implicit in the notion of combinatorial creativity is the idea that a combination is representationally valid; that is, not a mere hodgepodge of basic concepts, but an orderly combination that observes strong semantic criteria of the domain in question. In this paper we explore a model of combinatorial creativity in which combinations are organized via a semantic grammar. The term grammar seems particularly appropriate here, since it essentially represents a collection of templates into which elements can be inserted to form candidate (though not necessarily creative) combinations.

Viewing artefact generation in terms of lexical generation grants us a number of important advantages. Firstly, it permits the rapid acquisition of a large case-base of valid combinations. Secondly, it provides a simple means of validating lexical descriptions via web-search. This in turn allows us to reconsider Bodens notion of P-Creativity as a graded rather than binary phenomenon, since some combinations will be relatively common on the web, while others, while present, will be exceedingly rare; the former is clearly less P-Creative than the latter. Finally, it allows us to leverage this graded notion of P-Creativity to provide a computational realization of H-Creativity.

2 GASTRONAUT

Gastronaut is an architecture for exploring combinatorial creativity in the domain of food and culinary creation. We stress that the goal of Gastronaut is not the pursuit of culinary creativity for its own sake, but as a metaphor for combinatorial creativity in general. Our intent here is to determine the extent to which lexical creativity is a meaningful precursor to artefactual creativity.

2.1 A Semantic Grammar of Food

Cooking and chemistry are perhaps the prototypical application of combinatorial creativity. In each case, a subset of a known inventory of basic elements are combined to achieve novel and useful results. In both domains, combination is a controlled process, which must be done obeying domain constraints. This control is reflected in the way a combination is lexicalized, giving prominence to certain elements over others. To model the combinatorial structures, or templates, we employ a grammar built upon the following semantic categories :

1. <**meth**>: Cooking methods such as "grilled", "toasted", "baked" and "fried".
2. <**ing**>: Primitive ingredients such as "potato", "ginger" and "tomato".
3. <**dish**>: Complex items to which others are added, such as "pie", "pasta", "stew".

There are cases when a concept belongs to multiple categories, depending on the context. For instance, "beef" is both a primitive ingredient (as in beef pie) and a dish ingredient (as in ginger beef). These elements can be combined in the following ways:

<**ing**> <**dish**>*(e.g., cheese pizza)*
<**meth**> <**ing**> *(e.g., baked trout)*
<**ing**> with <**ing**> *(e.g., potatoes with cream)*
<**dish**> with <**ing**> *(e.g., pizza with anchovies)*

[1] School of Computer Science and Informatics, University College Dublin, Ireland, email: {Ioana.Butnariu, Tony.Veale}@UCD.ie

<**ing**> <**dish**> with <**ing**> *(e.g., cheese pizza with anchovies)*
<**ing**> with <**ing**> and <**ing**> *(e.g., beef with cheese and nuts)*
<**ing**> and <**ing**> <**dish**> *(e.g., beef and carrot stew)*
<**dish**> with <**ing**> and <**ing**> *(e.g., pizza with cheese and sausage)*

At present, Gastronaut works with an inventory of 7 cooking methods (<**meth**>), 60 basic ingredients (<**ing**>), and 25 complex ingredients (<**dish**>).

2.2 The Creative Cycle in Gastronaut

In creating new lexical combinations, and thus, new culinary combinations, Gastronaut employs three distinct sub-processes:

1. **Generation:** Gastronaut must fill in the appropriate templates with all possible, well-formed combinations of lexical ingredients.
2. **Retrieval:** The lexical combinations thus generated are sought on the world-wide-web using a search engine (such as AltaVista or Google); those combinations for which valid web-pages are retrieved are deemed to be "web validated" and a P-Creativity score, $P_{creative}$, is assigned (as described in section 3). It is important here to work not with raw web counts, but with counts adjusted for the effects of noise. Furthermore, Gastronaut does not work with absolute counts, but web-counts that are framed relative to specific, category-level expectations about the ingredients involved. Thus, a combination involving chocolate is assessed relative to the prototypical "chocolate" dish (e.g., "chocolate cake").
3. **Justification:** As befitting the nature of creativity, many genuinely useful combinations will not be found in a search of web-content. However, separating creative combinations from genuinely distasteful ones (mackerel ice-cream with prunes?) is the acid-test of a creative system. Since justification is central to the diagnosis of H-creativity, we consider this topic in greater depth in section 4. Gastronaut justifies such combinations in terms of known combinations. As such, the retrieval process of step 2 becomes, in effect, one of case-base construction.

3 ESTIMATING P-CREATIVITY AS A SCALAR QUANTITY

Boden [1,2] makes a key distinction between *H-Creativity* (in which a system produces an idea/artefact that has never been discovered by anyone before) and *P-Creativity* (in which a system produces an idea/artefact that has been already discovered by others, which is new to the system).

However, P-Creativity should not be seen as a simple binary distinction between original/unoriginal. Rather, since some lexical combinations will be more obvious than others (and thus cause a multitude of documents to be retrieved from the world-wide-web), P-Creativity should instead be considered a gradable phenomenon that is dependent on the cultural frequency of the combination. Combinations that retrieve very few documents are rare, and thus more novel, than combinations that retrieve very many.

3.1 Web Frequency as Typicality

One useful consequence of identifying ideas with their lexical expression is that the typicality of such ideas can be estimated via their frequency on the world-wide-web. Thus, we can estimate the P-creativity of the combination chocolate and pizza, denoted $P_{creative}$(chocolate:pizza), in terms of the web frequency of the query phrase "chocolate pizza". We refer to the raw page-count of a lexical combination X:Y as f_{raw}(X:Y). This statistic provides a noisy picture of the acceptability of a given combination, since many instances of the combination "X Y" will not be true occurrences of the idea X:Y. Rather, many instances will contain intervening punctuation that mark the instance as a chance co-occurrence of words rather than a valid expression of the desired combination. Since f_{raw} can produce page-counts in the millions, it is impractical to validate every instance. Thus, we validate the first 100 instances of every lexical combination, to estimate a probability P_{noise}(X:Y) that a given instance is invalid. In turn, this allows us to express f_{adj}(X:Y), the noise-adjusted frequency of "X Y", as:

$$f_{adj}(X:Y) = f_{raw}(X:Y)(1 - P_{noise}(X:Y))$$

3.2 Estimating P-Creativity

Intuitively, one expects a strong correlation between $f_{adj}(X:Y)^{-1}$ and the P-Creativity of the combination X:Y, since atypicality should derive from relative novelty. However, absolute web frequencies can be misleading [6], and what is needed is a way of framing the frequencies in the context of the appropriate categories. The novelty of a combination like "chocolate pizza" must be measured with respect to two categories: the prototypical chocolate dish (e.g., "chocolate cake"), and the prototypical pizza dish (e.g., "cheese pizza"). Let $C_{proto}(X)$ denote the prototypical combination in which X occurs, and $f_{adj}(C_{proto}(X))$ as its noise-adjusted web frequency; $C_{proto}(X)$ can be determined by identifying the element Y that maximizes the adjusted quantity $f_{adj}(X:Y)$. The relative atypicality of the combination X:Y, and thus its P-Creativity, can then be estimated as:

$$P_{creative}(X:Y) = 1 - (ln(f_{adj}(X:Y))/ln(f_{adj}(C_{proto}(X))) + ln(f_{adj}(X:Y))/ln(f_{adj}(C_{proto}(Y))))/2$$

For combinations of three elements, we employ piecewise evaluation, as every combination X:Y:Z necessitates three pairwise combinations X:Y, Y:Z and X:Z:

$$P_{creative}(X:Y:Z) = average(P_{creative}(X:Y), P_{creative}(Y:Z), P_{creative}(X:Z))$$

According to the semantic grammar, the three-way combination X:Y:Z can be realized syntactically in up to six forms, according to the category type of the concepts X, Y and Z. For instance, if Y and Z are simple ingredients, but X is a complex one than we have the following: "Y and Z X", "Z and Y X", "X with Y and Z", "X with Z and Y", "Z X with Y" or "Y X with Z". In turn, the combination X:Y, can be realized by the grammar up to two or three syntactical forms as "Y X", "X with Y" or "Y with X". The above formulae for f_{adj} and $P_{creative}$ employ the syntactic arrangements that minimizes the value of the resulting score. Three-way combinations are most creative then when they contrive three creative pairwise combinations. To the extent that one or more of these subordinate pairings are not creative, the overall trio of ingredients will be seen as less creative also.

3.3 An Example

Consider the creative combination chocolate:pizza, lexicalized by Gastronaut with "chocolate pizza" (undoubtedly a dessert

dish). For this combination, f_{raw}(chocolate:pizza) = 66,800, P_{noise}(chocolate:pizza) = 0, so f_{adj}(chocolate:pizza) = 66,800. Now, C_{proto}(chocolate) = chocolate:cake, and C_{proto}(pizza) = sausage:pizza, where f_{adj}(sausage:pizza) = 777,000 and f_{adj}(chocolate:cake) = 4,170,000. Thus, $P_{creative}$(chocolate:pizza) = 0.22. The value indicates that chocolate:pizza can be deemed unusual, at least compared with the more common chocolate:soufflé, with $P_{creative}$(chocolate:soufflé) = 0.09, or cheese:pizza, with $P_{creative}$(cheese:pizza) = 0.02.

Now, because $P_{creative}$(chocolate:peanut) = 0.25 and $P_{creative}$(peanut:pizza) = 0.63, Gastronaut assesses $P_{creative}$(chocolate:pizza:peanut) = 0.36 according to the formula for three-way combination. We place this combination among the P-Creative ones, as we found f_{raw}(chocolate:pizza:peanut) = 2 and f_{adj}(chocolate:pizza:peanut) = 1 for the lexical label "chocolate pizza with peanuts".

4 ESTIMATING H-CREATIVITY AS A SCALAR QUANTITY

A combination that does not occur on the world-wide-web may either be P-Creative (yet never indexed), H-Creative, or merely nonsensical. We explored the space of pairwise combinations obtained in the generation process and stored in a case-base those for which we were able to asses a web-valid lexical label according to our grammar templates. For the wider space of three-way combinations, the retrieval process is not sufficient: we need a non-corpus approach to discriminate among combinations that cannot be validated via the web.

Fortunately, Gastronaut can exploit the fact that a novel combination should exhibit a coherence with a body of known P-Creative combinations (as stored in the accumulated case-base of known/retrieved P-Creative combinations) that collectively suggest it to be sensible. Gastronaut thus needs to determine just how coherent a novel combination is with respect to the set of existing valid combinations. One approach is to decompose combinations into smaller combinations that are "chained" together. If these smaller combinations can all be validated as P-Creative, then the whole combination can be seen as sufficiently credible to be deemed H-Creative.

That is, Gastronaut can justify novel combinations of the form X:Y:Z by demonstrating the affinity between the concepts in all pairwise combinations X:Y, Y:Z and X:Z, as follows:

$$\text{X:Y:Z} \leftarrow justifies \text{ - X:Y} \wedge \text{Y:Z} \wedge \text{X:Z}$$

In effect, this decompositional analysis demonstrates that each ingredient is compatible with all others, thereby suggesting the combination of all three together to be valid. Note that we use the form X:Y:Z to denote any lexical phrasing of the combination of X with Y and Z, and X:Y to denote any lexical phrasing of the combination of X with Y, according to the templates in our grammar. That is, this scheme generalizes over all possible syntactic instantiations of the combination: if any combination of X and Y is known to have a non-zero $P_{creative}$ score, then X:Y is said to be a valid combination that partially justifies X:Y:Z.

This further suggests that H-Creativity can be assessed in the same way as P-Creativity; the difference, of course, is that P-creative combinations are web-validated, while H-Creative are not (and can not be) web-validated. In other respects, the scale for each type of creativity is parallel:

$$H_{creative}(X:Y:Z) = average(P_{creative}(X:Y), P_{creative}(Y:Z), P_{creative}(X:Z))$$

where f_{adj} (X:Y) > 0, f_{adj} (Y:Z) > 0 and f_{adj} (X:Z) > 0.

A H-Creative three-way combination X:Y:Z is a combination for which we couldn't find a valid syntactical representation on the web. Gastronaut can propose such a lexical form, among all the possible ones in our grammar, that is suggested by X:Y, X:Z and Y:Z subcomponents. That is possible, as the templates for X:Y:Z combinations in our semantic grammar can be decomposed into templates for pairwise combinations.

4.1 An Example

We consider the novel combination chocolate:pizza:carrot of type X:Y:Z. This combination seems credible if we can show that three other combinations are valid:

X:Y (e.g. chocolate:pizza)
Y:Z (e.g. chocolate:carrot)
X:Z (e.g. pizza:carrot)

We obtain $P_{creative}$(chocolate:pizza:carrot) = 0.46, as the average of $P_{creative}$(chocolate:pizza) = 0.22, $P_{creative}$(carrot:pizza) = 0.64 and $P_{creative}$(carrot:chocolate) = 0.54.

None of the lexical forms for chocolate:pizza:carrot generated by the grammar templates are web-valid. After a close analysis of the syntactical forms for the subcomponents "chocolate pizza", "carrot pizza" and "carrot chocolate", Gastronaut picks "carrot and chocolate pizza" as the most appropriate syntactical form for this combination.

5 EMPIRICAL EVALUATION

In order to measure the creativity involved in the combinations of culinary ingredients, we used the formulae described in section 3.2. The results show a good correlation (-0.4) between f_{adj}(X:Y) and $P_{creative}$(X:Y). If we extend the same formula based on web frequencies for three-ways combinations X:Y:Z, we observe an insignificant correlation of -0.02 between f_{adj}(X:Y:Z) and $P_{creative}$(X:Y:Z); as a result we asses creativity for the more complex X:Y:Z combinations in terms of creativity values of pairwise combinations.

Table 1 presents a breakdown, by combination type, of the number of lexical labels generated with Gastronaut semantic grammar, and the percent of web-valid ones:

Table 1. Number of lexical instances generated by each template and the percent of web valid instances.

Template	# instances	web valid
<**ing**> <**dish**>	1,329	1,266 (95%)
<**meth**> <**ing**>	360	324 (90%)
<**ing**> with <**ing**>	2,679	1,455 (54%)
<**dish**> with <**ing**>	1,329	1,036 (78%)
<**ing**> <**dish**> with <**ing**>	23,667	1164 (5%)
<**ing**> with <**ing**> and <**ing**>	56,044	643 (1.2%)
<**ing**> and <**ing**> <**dish**>	27,961	2365 (8.5%)
<**dish**> with <**ing**> and <**ing**>	20,763	352 (1.6%)

The results show that lexical instances for pairwise combinations have a good web coverage, in contrast with the more complex three-way combinations. In consequence, we used a non-corpus approach in assessing creativity to three-way combinations.

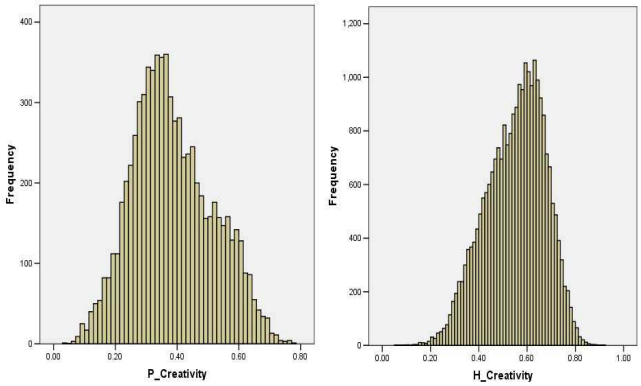

Figure 1. P-Creativity and H-Creativity for three-way combinations.

We found that the mean P-Creativity score for all web-validated three-way combinations is 0.38, with a variance of 0.016, while the mean H-Creativity score for all three-way combinations (not web-validated) is 0.55, with a variance of 0.0014. The histograms of the distribution of P-Creativity and H-Creativity scores are shown in Figure 1.

Computational creativity in food domain is a fruitful process. The empirical results show that 80% of X:Y:Z combinations generated by Gastronaut are potentially innovative products, as these are not web-validated. Our system ranks these H-Creative candidates with high creativity scores (an average of 0.55), while for most of the P-Creativity combinations, the scores remain within 0-0.5 range. We conclude that Gastronaut's measurement for the degree of novelty involved in each combination is coherent with and partially justified by the data retrieved from the world-wide- web.

6 CONCLUSIONS

P-Creativity and H-Creativity are not binary notions, but scalar quantities that vary along a continuum. In this paper, we have attempted to numerically model this continuity using web-based document counts of the lexical realizations of a given combination. Pairwise combinations of elements form the bedrock of our numerical approach, allowing P-Creativity and H-Creativity scores for three-way combinations to be assessed using the same formula. The difference, of course, is that P-Creative combinations are validated via web-search, while H-Creative combinations cannot be found on the web. However, as such combinations do begin to appear on the web, and shift from H-Creative to P-Creative status, their creativity scores will not experience a discontinuity. The creativity of three-way combinations is not assessed in terms of the web-frequency of the corresponding three-element phrases, but in terms of the relative adjusted web-frequency of the pairwise combinations that each implies (this decision is borne out by the poor correlation of -0.02 between web-frequency and P-creativity for three way combinations). Creativity scores for three-way combinations are thus stable as these combinations eventually transition from P- to H-creativity status.

Assessing appropriate lexical forms for a given combination is a key issue in Gastronaut. We have experienced that combinations formed with more than two concepts are difficult to be lexicalized following restrictive patterns. For instance, with our present semantic grammar none of the labels for the combination potato:tomato:sausage is web-validated. When we make a search on Google, we can find appropriate lexical expressions like "sausage with mashed potato and grilled tomato". In consequence, we intend to extend our semantic grammar templates to be able to generate complex expressions as the one in the example above.

The Gastronaut system has also been used in developing an educational game, "Dr. FrankenFoods", in which a human user must attempt to construct creative culinary combinations from a given stock of ingredients. Creative games like these, intended to foster creative skills in children, require a computational system that is capable of recognizing not just P-Creativity, but H-Creativity. Though Dr. Frankenfoods is a rather simple application of creative computation, a new generation of creativity-enhancing games awaits developments of this kind.

ACKNOWLEDGEMENTS

We would like to thank the Enterprise Ireland for supporting this research through a grant from the Commercialization Fund.

REFERENCES

[1] M. Boden, *The Creative Mind: Myths and Mechanisms*, Weidenfeld and Nicholson, London, 1990.
[2] M. Boden, Computational models of creativity, *Handbook of Creativity*, 351-373, 1999.
[3] K. J. Hammond, CHEF: A model of case-based planning, *In Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86)*, Philadelphia, PA, 1986.
[4] J. Hayes, T.Veale, and N. Seco, Creative discovery in the lexical validation gap, *Journal of Computer speech and Language*, Volume 19, Issue 4, 513–523, , 2005.
[5] T.R. Hinrichs, Problem Solving in Open Words: A case study in design, *Hillsdale, NJ: Lawrence Erlbaum Associates*, 1995.
[6] F. Keller, M. Lapata, Using the web to obtain frequencies for unseen bigrams, *Computational Linguistics*, 2003.
[7] F. Pereira, The process of creative thinking, Ph.D. thesis, 2004.
[8] G. Ritchie, Assessing creativity, *In Proceedings of the AISB Symposium on Artificial Intelligence and Creativity in Arts and Science, 3–11, York, England*, 2001.
[9] T. Veale, Creative Information Retrieval, *In the proceedings of CICLing*, ed. Alexander Gelbukh, Seoul, 2004.
[10] G. Wiggins,Categorizing Creative Systems, *In the proceedings of the 3rd Workshop on Creative Systems, IJCAI03*, Acapulco, Mexico, 2003.