

Style Variation in Cooking Recipes

Jing Lin¹ and Chris Mellish² and Ehud Reiter³

Abstract. Human text by specific authors has varied characteristics. Even when different authors write about identical topics within a fixed genre under a predefined domain, their text will still be different. For example, in food recipes there are style variations between authors even when describing the same cooking actions and foods. This paper summarises style variations in food recipes written by different people.

1 Introduction

Nowadays it is common to use corpora of human-authored texts to acquire knowledge in natural language generation (NLG), such as STOP [10] and SumTime [12], and so on. To make sure the corpora are big enough and broad enough, the corpora generally contain texts from several authors. When corpora are analysed, the differences between authors are normally ignored. Reiter and Sripada [9, 8] have suggested that “There are substantial variations between individual writers, which reduces the effectiveness of corpus-based learning”. Our ultimate goal is to build an NLG system which translates recipes written by one author into the style of another author. This requires a good model of what aspects of recipes differ between authors; the corpus analysis presented here is a step towards such a model.

In the NLG area, some researchers have investigated producing text with varied styles. Hovy [4, 5] discussed rhetorical goals that are “the existence of a level of organization mediating between communicative goals and generator decisions”. Furthermore, Stamatas et al. [13] can produce outputs based on different user requirements, such as written style, tone, and so on. None of these researches focuses on generating texts with the style from an author. In literary comparison, *stylometry*, which is a subcategory of *attributing authorship*, researchers try to identify an author’s style by building a set of measurable patterns [14]. The purpose of *stylometry* is to identify the author of texts, instead of describing the writing characteristics. Most of the defined *style markers* of stylometry are not enough to identify writing characteristics. For example, Burrows [1] only considers the frequencies of a group of the top (typically 30 or 50) frequent words. Moreover, some researchers used character-level n-grams recently, such as Kešelj [6], Peng [7], and Clement and Sharp [2], to consider the impact of morphological variations of words.

We introduce our corpora based on authors in Section 2 and give two example recipes by different authors in Section 3. Section 4 describes style variations in food recipes, which are categorised in at the sentence level and at the recipe level. The final section is our conclusion in Section 5, which relates this work to the problem of generating texts with styles of authors.

¹ University of Aberdeen, UK, email: j.lin@abdn.ac.uk

² University of Aberdeen, UK, email: c.mellish@abdn.ac.uk

³ University of Aberdeen, UK, email: e.reiter@abdn.ac.uk

2 Our Corpora

To catch style variations between authors, we need to collect good quality corpora. Certain genres are less suited to analysing writing differences, such as daily news and fictions. Burrow [1] suggests that “in fiction: The language of its dialogue and that of its narrative usually differ from each other in some obvious and many less obvious ways”. The authors of daily news often narrate matters with quotes from involved people and daily news is normally subjective from its authors, which make the text varied. Texts in food recipes, on the other hand, have features which are narrative, imperative, and objective. Although these features partly limit text variations, they also highlight styles between authors.

Based on the recipe domain, we collected a set of corpora based on authors from the Internet and recipe books. Table 1 presents information about our corpora. Our corpora include two recipe books, *Recipes for Health Eating* [3] and *Food for Health* [11]. There is an amateur recipe author: CM from the website (www.cooks.com). We also collected a few corpora of professional cooks from the BBC recipe website (www.bbc.co.uk/food/recipes).

Table 1. Author corpora information

Author Corpus	Corpus Information		
	Number of Recipes	Total Lines	Total Words
Recipes for Health Eating Dewhurst and Lockie [3]	76	961	9212
Food for Health Roe [11]	113	1347	11791
CM CM (www.cooks.com)	48	537	6432
Ainsley Harriott Harriott (www.bbc.co.uk)	87	891	14217
Antony Worrall Thompson Thompson (www.bbc.co.uk)	1026	7790	99791
James Martin Martin (www.bbc.co.uk)	1103	8996	119582
Hugh Fearnley-Whittingstall Fearnley-Whittingstall (www.bbc.co.uk)	77	932	12187

3 Two Example Recipes

We list two recipes as examples from our corpora to provide an idea what our recipes look like. Both of the recipes are fish pie recipes, but

they show many differences between one and the other. For example, Step 1 of Recipe One is delayed to Step 7 of Recipe Two. Since each cook intends to create recipes with his speciality, it is unlikely to find two recipes describing exactly the same dish. In the following examples, we omit the ingredient section of the recipes as these do not typically contain syntactically valid English sentences.

Recipe One Fish Pie (Dewhurst and Lockie, [3])

- 1 Switch on oven to 220C, 425F or Gas Mark 7 and grease a 1/2 litre ovenproof serving dish.
- 2 Collect ingredients.
- 3 Wash and peel potatoes and cut into evenly-sized pieces. Place in boiling water and cook for approximately 20 minutes until soft.
- 4 Drain and mash potatoes when cooked. Add low fat spread and stir thoroughly.
- 5 Meanwhile, rinse and dry fish and poach gently in milk until fish flakes easily (approximately 10-15 minutes).
- 6 Drain and flake fish. Retain milk for sauce.
- 7 Measure warm milk and if necessary make up to 75 ml with water.
- 8 Mix mashed potatoes, flaked fish and milk.
- 9 Wash and finely chop parsley and add to fish and potato mixture with a good pinch of grated nutmeg.
- 10 Place mixture in greased ovenproof serving dish and using a fork, mark a pattern on top.
- 11 Bake for 10-15 minutes until brown on top.
- 12 Wash tomatoes, cut into wedges and arrange neatly on top. Place lemon slice and parsley in middle.

Recipe Two Creamy Fish Pie (Fearnley-Whittingstall, BBC online recipes)

- 1 Place the fillets of fish in a medium saucepan. Add the milk, onion, carrot, celery, bay leaf, a couple of stalks of parsley and the peppercorns.
- 2 Place the pan on a low heat and let the milk heat up gently. As soon as it comes to a simmer, switch off the heat and cover the pan. The fish will continue cooking in the hot milk.
- 3 Meanwhile, peel the potatoes, cut them into even, bite-sized chunks and put them in a large pan. Add just enough water to cover and put the pan on the hob over a high heat. Add a teaspoon of salt and let the water come to the boil. Lower the heat to a simmer and cook the potatoes until they are just tender.
- 4 Carefully drain the potatoes and allow them to cool in a colander for a minute or two. Return them to the pan and mash them, adding 50g/2oz of the butter, cut into cubes.
- 5 Stand a sieve over a large jug and tip in the fish and milk mixture. Wash the pan in which the fish was cooked and dry it well.
- 6 Add 3-4 tablespoons of the fishy milk to the mash and stir it in well. Add some freshly ground black pepper, taste the mash, and add some salt if you think it needs it. Put the mash to one side.
- 7 Heat the oven to 200C/400F/Gas 6.
- 8 To make the bechamel sauce, put the remaining 75g/2 1/2oz butter in the clean pan and melt it over a low to medium heat.

Add the flour and stir well with a wooden spoon to make a roux. Cook for two minutes, stirring every few seconds. Then gently whisk in one third of the hot fishy milk. The paste will quickly turn into a very thick sauce. Add another third of the milk, whisking all the time, and then the final third, so you end up with a creamy sauce. Season the bechamel with salt and freshly ground black pepper, turn the heat down to very low, and let the sauce bubble gently for five minutes while you prepare the fish.

- 9 Remove the vegetables, herbs and peppercorns from the fish and discard. Carefully pick up a chunk of fish. Peel off any skin and discard, then gently feel the flesh between your fingers for bones, being careful not to over-shred the fish. Put the boneless fish on a clean plate.
- 10 Turn off the heat under the bechamel and add the fish to the sauce. Add the prawns, then chop the remaining parsley and stir this in too. Taste the sauce once more and add more seasoning, to taste.
- 11 Generously butter a pie dish and pour in the fishy bechamel. Spoon over the mash and spread it carefully across the surface of the fish sauce. Dot a little extra butter over the top of the pie.
- 12 Wearing oven gloves, put the pie in the oven and bake for about 25 minutes or until the top is starting to brown and the fishy sauce is bubbling up the sides of the mash.
- 13 Serve with buttered minted peas and crusty bread to mop up the sauce.

4 Style Variations

There are different types of style features with examples in food recipes. We categorised them into sentence level and recipe level features. The style features at the sentence level are those that can be identified within one sentence, and do not have many connections outside of their sentence. The style features at the recipe level are those features which involve relationships between sentences, or cannot be simply defined within one sentence.

4.1 Sentence Level Features

Sentence level features include lexical preferences (Section 4.1.1), skipping objects (Section 4.1.2), content selection (Section 4.1.3), other differences at the sentence level (Section 4.1.4), and orthographical differences (spelling) (Section 4.1.5).

4.1.1 Lexical Preferences

Authors express the same action with different words, which can be synonyms with each other, or can be not. For example in Recipe One Step 1, Dewhurst and Lockie say ‘*Switch on oven*’ whereas in Recipe Two Step 7, Fearnley-Whittingstall says ‘*Heat the oven*’. The following other examples show similar variation in the same action from other authors.

Example 1 “*Put the oven on.*” (Roe, [11])

Example 2 “*Preheat the oven to 200C/400F/Gas6.*” (Thompson, BBC online recipes)

In the above four examples, these different verbs reflect the fact that the authors have different lexical preferences to express the same action — preheating the oven. Table 2 shows the frequency of the 4

Table 2. Sentence and word information of corpora of different authors

Authors	Following Objects (oven or grill)			
	Switch on	Put on	Preheat	Heat
Dewhurst and Lockie	35	2	0	0
Roe	0	42	0	0
Thompson	0	0	298	2
Fearnley-Whittingstall	0	0	4	7

verbs by the authors given objects ‘oven’ or ‘grill’. Since ‘switch on’ only appears in the authors, Dewhurst and Lockie, it can be considered as idiolect. In the recipe domain, idiolects may cover verbs, and nouns as well. For instance, some authors decide to use ‘frypan’ instead of ‘pan’. Furthermore, Dewhurst and Lockie and Roe never use the word, ‘preheat’, in their recipes. Fearnley-Whittingstall’s recipes are various, including desserts, main courses, and so on. Only small portion of his recipes involves bakeware, ‘oven’ or ‘grill’.

A similar situation happens for other verbs, as in the following examples 3 and 4.

Example 3 “*Extract* juice from orange and add this with the water to the saucepan.” ([3])

Example 4 “Finely grate the ginger and *squeeze* out the juice into a shallow non-metallic dish.” (Harriott, BBC online Recipes)

4.1.2 Skipping Objects

Some authors prefer to skip the objects of their verbs if they have been mentioned before, whereas others always include objects. See the following examples.

Example 5 “Blanch the sweet *potato* in boiling water for 2-3 minutes until softened. Transfer to a bowl and *mash* with the cream and herbs.” (Thompson, BBC online recipes)

Example 6 “Wash and peel *potatoes* and cut into evenly-sized pieces. Place in boiling water and cook for approximately 20 minutes until soft. Drain and *mash potatoes* when cooked.” (Step3 & Step4 of Recipe One) (Dewhurst and Lockie, [3])

Thompson always skips the object, ‘potatoes’, when he mentions the action, ‘mash’. Furthermore Fearnley-Whittingstall includes the object as a pronoun, ‘them’, in Step 4 in Recipe Two.

4.1.3 Content Selection

Authors present conditions using different information.

Example 7 “Bake until *golden*.” (Roe, [11])

In the above example, in Step 11 of Recipe One, and in Step 12 of Recipe Two, authors describe the appearance of the food when it is cooked, by using ‘golden’ or ‘brown’. Both Dewhurst and Lockie and Fearnley-Whittingstall also describe the cooking time in their baking procedure. Fearnley-Whittingstall adds the delicious detail that ‘the fishy sauce is bubbling up’ in contrast to the simpler descriptions by Roe and Dewhurst and Lockie. Roe uses subjective descriptions, such as ‘golden’, ‘tender’, and ‘firm’ more often, instead of time information in her recipes. Her ways are more suitable for experienced cooks.

Some authors provide less details or no details, but others prefer to introduce actions with clear process descriptions. This happens especially in using certain verbs, like ‘drain’ in the following examples.

Example 8 “Drain.” (Roe, [11])

Example 9 “Drain.” (Dewhurst and Lockie, [3])

Example 10 “Drain the beef from the marinade into a colander over a glass bowl.” (Thompson, BBC online recipes)

When the authors present the action, ‘drain’, Roe and Dewhurst and Lockie rarely include further information, but Thompson most of the time presents detailed information related to the action, ‘drain’. This is in contrast to the example of ‘mash’ in the previous section, in which Dewhurst and Lockie gave more information than Thompson.

4.1.4 Other Differences at the Sentence Level

Because of the freedom of the English language, authors can structure the same words into different sequences.

Example 11 “Peel and *finely* chop onion.” (Dewhurst and Lockie, [3])

Example 12 “Peel and chop the onion *finely*.” (Roe, [11])

There are some unusual grammatical presentations in our corpora, for instance Dewhurst and Lockie always say ‘Switch on oven’ in Step 1 of Recipe One, but never say ‘Switch on the oven’.

4.1.5 Orthographical Differences

There are some orthographical differences, which are recognised differently in computer systems. However, they should be considered as one word presenting in different forms, not two different words.

Example 13 “Pre-heat the oven to 220C/425F/Gas 7.” (Thompson, BBC online recipes)

Thompson uses the word, ‘preheat’, in the form, ‘pre-heat’, 17 times. Moreover, Dewhurst and Lockie use ‘Gas Mark 7’ in Step 1 of Recipe One instead Thompson just uses ‘Gas 7’ in the previous example.

4.2 Recipe Level Features

At the recipe level, we find content differences (Section 4.2.1), order differences (Section 4.2.2), structure differences (Section 4.2.3), and aggregation (Section 4.2.4).

4.2.1 Content Differences

Some authors, such as Dewhurst and Lockie, always present certain cooking actions, but others, such as Thompson, skip these actions in recipe descriptions. Thompson never mentions ‘collect ingredients’ and ‘wash potatoes’ in the cooking method section, but Dewhurst and Lockie always does as showing in Step 2 of Recipe One.

On the other hand, most cooks only describe the food making process. However, Fearnley-Whittingstall also involves other actions. For example, he mentioned ‘wearing oven gloves’ in Step 12 of Recipe Two and ‘washing the pan’ in Step 5 of Recipe Two.

4.2.2 Order Differences

Different authors describe their recipes using a different order. In the Dewhurst and Lockie corpus ingredients are prepared when the time comes in the cooking process, so that the actions, ‘wash and peel potatoes’, are described in the cooking methods. Fearnley-Whittingstall follows this idea as well. On the other hand, Thompson assumes all ingredients should be prepared properly before cooking starts. For instance, ‘1/2 sweet potato, peeled and diced’ in the ingredient section, instead of as Step 3 of Recipe One in the method part.

Describing preparing ingredients in the cooking methods is more suitable for experienced cooking learners, since they could handle many cooking processes at the same time. For beginners, Thompson’s way could save a lot of fuss.

4.2.3 Logic Preferences

In Section 3, we presented two recipes for fish pie, which contain common cooking actions. Table 3 shows the sequence of these actions appearing in the recipes. Since these two recipes are quite similar, it shows that the two authors have different logic preferences.

Table 3. Common cooking actions between two fish recipes

Action Code	Cooking Action
Action <i>a</i>	preheating the oven
Action <i>b</i>	making mash potatoes
Action <i>c</i>	cooking fish
Action <i>d</i>	baking the pie
Authors	Action Sequence
Dewhurst and Lockie	$a \rightarrow b \rightarrow c \rightarrow d$
Fearnley-Whittingstall	$c \rightarrow b \rightarrow a \rightarrow d$

4.2.4 Aggregation

Sometimes, one sentence in a cooking recipe contains two actions or more. This is called action aggregation. We expected to find different action aggregation habits from different authors. To our surprise, the action aggregations of different authors are unpredictable in our corpora. For instance, the authors, Dewhurst and Lockie, have aggregated the actions, ‘switch on’ and ‘grease’, 10 times as in Step 1 of Recipe One, and have not aggregated them 20 times as in example 16. The author, Martin, also aggregated the actions, ‘preheat’ and ‘grease’, 4 times as in example 17 out of 15 times. Since action aggregation is a common feature in food recipes, though it has no

uniform style for each author, we feel it is worth mentioning here. Table 4 shows all the aggregation information about this section’s examples.

Example 14 “Switch on oven to 200C, 400F or Gas Mark 6 and collect ingredients.” (Dewhurst and Lockie, [3])

Example 15 “Switch on oven to 180C, 350F or Gas Mark 4. Collect ingredients.” (Dewhurst and Lockie, [3])

Example 16 “Switch on oven to 190C, 375F or Gas Mark 5. Grease a 1/ litre ovenproof serving dish.” (Dewhurst and Lockie, [3])

Example 17 “Preheat the oven to 170C/325F/Gas 3 and grease a muffin tray.” (Martin, BBC online recipes)

Example 18 “Preheat the oven to 200C/400F/Gas 6. Grease a baking dish with butter and dust with flour.” (Martin, BBC online recipes)

5 Conclusion

In this paper, we collected and analysed many style variations between different authors. Some style features are at the sentence level. For example, most authors in our corpora have their own preferred words when they describe some actions in lexical preferences. Other style features locate at the recipe level, which rely on the subjective decision of an author. Logic preferences, for instance, clearly show the logic of an author.

In some style features, authors show clear preferences, as Dewhurst and Lockie always use ‘switch on the oven’ and Martin uses ‘preheat the oven’. However, the aggregation between different authors are unpredictable. Another interesting observation is that authors include varied levels of details in their recipes, as Fearnley-Whittingstall even describes ‘bubbling up’ in the baking step of Recipe Two.

Before a NLG system is capable of generating text with these writing style features, these style knowledge need to be extracted from corpus first in the knowledge acquisition process. Some style features require information from other style features. For example, all writing style features at the recipe level need information supports from lexical preferences and skipping objects. Therefore, to analyse the subtle style features, lexical preferences and skipping objects are the starting point in the knowledge acquisition. Only if all actions have been properly identified through lexical preferences, logic preferences, for example, can be identified.

Style features have their implications for NLG at different levels, if a NLG system intends to generate text with all these features. In the content determination, content selection at the recipe level and logic preferences need to be considered. Content selection at the sentence level and other differences involve in the sentence determination. Finally, lexical preferences, skip objects, and orthographical differences influence in lexical choice.

REFERENCES

- [1] John F. Burrow, ‘Word-patterns and story-shapes: the statistical analysis of narrative style’, *Literary and Linguistic Computing*, vol 2(2), pp 61–70, (1987).
- [2] R. Clement and D. Sharp, ‘Ngram and bayesian classification of documents for topic and authorship’, *Literary and Linguistic Computing*, vol 18(4), pp 423–447, (2003).
- [3] Y. Dewhurst and G. Lockie, *Recipes for Healthy Eating*, Heinemann Educational Books, 1986.
- [4] E. H. Hovy, *Generating Natural Language Under Pragmatic Constraints*, Lawrence Erlbaum, Hillsdale, NJ., 1988.

Table 4. Action aggregation Information about example 15 to example 19

Switch on — collect		Switch on — grease		Preheat — grease	
Aggregated	Separated	Aggregated	Separated	Aggregated	Separated
2	5	10	20	4	12
Total 7		Total 30		Total 16	

- [5] E. H. Hovy, 'Pragmatics and natural language generation', *Artificial Intelligence*, vol 43, pp 153–197, (1990).
- [6] V. Kešelj, F. C. Peng, N. Cercone, and C. Thomas, 'N-gram-based author profiles for authorship attribution', *Proceedings the Pacific Association for Computational Linguistics*, (2003).
- [7] F. C. Peng, 'Language independent authorship attribution using character level language models', *Proceedings of the European Association for Computational Linguistics (EACL-03)*, (2003).
- [8] E. Reiter and S. Sripada, 'Human variation and lexical choice', *Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL-02)*, (2002).
- [9] E. Reiter and S. Sripada, 'Should corpora texts be gold standards for nlg?', *Proceedings of the 2nd International Conference on Natural Language Generation (INLG-02)*, pp 97–104, (2002).
- [10] Robertson R. Reiter, E. and Liesl Osman, 'Knowledge acquisition for natural language generation', *Proceedings of the First International Conference on Natural Language Generation (INLG-00)*, pp 207–215.
- [11] D. Roe, *Food for Health*, Longman, 1990.
- [12] Reiter E. Hunter J. Yu J. Sripada, S. and I. Davy.
- [13] E. Stamatatos, S. Michos, N. Fakotakis, and G. Kokkinakis, 'A user-assisted business letter generator dealing with text's stylistic variations', *the Ninth International Conference on Tools with Artificial Intelligence (TAI-97)*, (1997).
- [14] N. Fakotakis Stamatatos, E. and G. Kokkinakis., 'Computer-based authorship attribution without lexical measures', *Computers and the Humanities*, (2001).