

As American as Apple Pie: A Search to Define Americanness in the Context of all Apple Pie Recipes on the Internet

Authors: Matthew A. Charleston¹, Joshua M. Shaw², Lennon J. Anderson³

Affiliations: 1. Former President Ohio State Bread Club
2. Top 5% Hearthstone Player
3. Former President Ohio State Hacking Club

Abstract

Since the halcyon days of Johnny Appleseed, America has been strongly associated with apples and more specifically apple pie. For what can more embody the American spirit than the sticky, sugary, (high fructose corn syrupy?), buttery, and flaky masterpiece that is the apple pie. However, neither the apple nor the apple pie originate in America. The apple being the oldest fruit (God, 1400 BC) and the pie one of the oldest desserts. In order to best embrace, define, restrict, and confine the American spirit, it is necessary to define the concept of Americanness, and what better local to scour than the venerable apple pie.

In this paper we condition and filter the Recipe NLG dataset into a list of ingredients associated with each apple pie recipe. We then sort the recipes into American and un-American Apple pies and calculate, based on the mean quantity of ingredients, what makes an American and un-American apple pie. Finally, the average American Pie is baked for gustatory analysis to determine if the average American has good taste.

Introduction

The common apple is the most ubiquitous and culturally important fruit crop in temperate climates (Cornille, 2012) and has been slowly domesticated into a household staple over thousands of years. Apples originated in prehistoric times and were omnipresent during the founding of America, developing simultaneously with the early colonies. Early Records show the apple tree growing in New England just after the founding of the Plymouth Colony. The spread of apples in America preceded the colonists, being distributed by the Native American Tribes who were then followed by the eccentric John Chapman (Appleseed). Starting around 1850 the apple began to transition from domestic orchard growth to important commercial proportions, ultimately becoming the most important fruit in existence, especially in the United States, the principle apple producing country in the world (Folger & Thomson, 1923).

As surely as the apple is enshrined in the hearts and history of Americans, so too is the apple pie. In fact, the apple pie is the only food officially recognized by the Department of Defense as an American Symbol (Bunschoten, 2014). The apple pie was a common item in the lives of the English colonists as they settled North America, and the

development of the apple pie mirrored the development of the growing nation. The first American cookbook, *American Cookery* (Simmons, 1796), contained the first recipe for an American apple pie. As the nation expanded westward and apple emissaries like Chapman spread the good news, corporate interests became involved as well. The Ohio Company offered land deeds to settlers who among other things, “set out at least 50 apple or pear trees” ensuring the spread of the delicious fruit. Despite a few brief deviant ventures like the removal of the upper crust in 1846 and the “Apple Pie without Apples” developed during the blockade of the confederate states (it substituted crackers and cream of tartar for apples), the pie spread through the breadth of the nation. The apple pie’s true association with national identity likely began during The Great War as soldiers wished longingly from the tranches for flavors of home. Through World War II as well, Betty Crocker and Elizabeth Roosevelt strode hand in hand down the pie aisle of America’s hearts, preaching home cooking and rationing frugality. In the post war-years the popularity of apple pie soared, becoming the favorite dessert of young Americans by 1958 (Bunschoten, 2014). As surely as the apple pie is distinctly American, it is equally unsure what separates the American Apple Pie from its cultural ancestors and rivals.

The current state of the art in recipe analysis has a strong basis in machine learning models. Recent models have been trained on large datasets of recipes to culturally adapt recipes between languages (Cao, et al., 2024) and even strategically substitute ingredients in recipes (Fatemi, et al., 2023).

One of the largest and most frequently cited collections of online recipe datasets is the Recipe1M+ dataset (Marín, et al., 2019) (Salvador, et al., 2017). This dataset is well known and has been used to study everything from recipe generation from images (Salvador, et al., 2019), to image generation from recipes (Salvador, et al., 2017). However, the Recipe1M+ dataset is not quite as public as described. The provided download link from Marin et.al. returned a http 500 error and told us to email the server administrator Yusuf for help. Unfortunately, despite our desperate and repeated attempts to contact Yusuf, our emails received no reply.

We next turned to the RecipeNLG dataset (Bien, et al., 2020), created at Polant’s Poznan University of Technology. Recipe NLG attempts to address the growing demand for recipe datasets to be used in deep learning experiments and especially ones centered on Natural Language Processing (NLP) tasks. The dataset contains the ingredients and respective quantities for over 2 million recipes, making it the largest recipe dataset at the time of its creation.

Experimental Setup

In keeping with the same spirit of rationalism that inspired America’s founding fathers, the decision was made to utilize human intelligence and mathematics to distill the Americanness from the Apple pie. The basic theory is laid out in Equation 1, where two exclusive subsets can be separated from the set of all Apple Pie recipes.

$$uAAP = AP - AAP \quad (1)$$

Symbol	Definition
--------	------------

AP	Apple Pie (Set of all Recipes)
AAP	American Apple Pie
uAAP	Un-American Apple Pie

Each subset is arranged as an array, where each row is a specific recipe and each column a specific ingredient. An example array is shown in Table 1. The ingredient columns of each subset can then be averaged to determine the quantity of ingredients in an average American Apple Pie and an average un-American Apple Pie. The differences in the average quantities of each ingredient are calculated according to Equation 2, to distinguish what makes an American Apple Pie unique, illustrated for example purposes in Table 2.

$$Americanness = \text{Mean}(\text{AAP}) - \text{Mean}(\text{uAAP}) \quad (2)$$

Table 1. Example Apple Pie Array

Recipe	Apples	Flour (c.)	Cinnamon (tsp.)	Sugar (c.)
Elaine's Apple Pie	6	2	3	1
George's Apple Pie	5	1	2	2
Jerry's American Apple Pie	6	1	4	1.5

Table 2. Example Mean Pie Arrays

Recipe	Apples	Flour (c.)	Cinnamon (tsp.)	Sugar (c.)
Mean(uAAP)	5.5	1.5	2.5	1.5
Mean(AAP)	6	1	4	1.5
Americanness	0.5	-0.5	1.5	0

In the example dataset here we can see that the largest distinguishing characteristic of Americanness is excess cinnamon with a slight preference for more apples and less flour.

Data Analysis

The Recipe NLG dataset was initially filtered to select only recipes that contained the words "Apple Pie", creating the new RecipeAP (apple pie) dataset with 6,803 apple pie recipes and 14,788 ingredients. Unfortunately for our heroes, the Recipe AP dataset, while being sufficiently conditioned for machine learning research, was ill-conditioned for straightforward mathematical analysis. Many ingredients are repeated with minor differences like "Pie Crust" vs "Pie Shell" or Major Differences like "Pie Crust" and "The dough may be made ahead of time and stored in the refrigerator for up to 5 days or in the freezer for up to two weeks. The pie is best served the same day.", or even completely absurd ingredients like "Dipping Fork". In order to analyze a concise and representative dataset, similar ingredient columns needed to be merged and erroneous recipes like "Bourbon Apple Pie Ice Cream" removed.

Further find and replace operations were conducted to standardize all measurement descriptions, replacing "c." with "cups" and "Tablespoons" with "Tbs.". Column merging was initially attempted manually using Microsoft Excel. Redundant columns were identified, then the entire array was sorted around redundant column 2 from high to low. The values from column 2 were cut and pasted into column 1 and the redundant column was deleted. This task was attempted for approximately 5 hours with the merging of 150 columns. This method was abandoned after it was calculated with the merge rate of 30 columns per hour, it would take 486 hours, or 20 full days to reduce the dataset to a reasonable 200 columns.

Returning to the RecipeAP dataset, while it is well suited to those working in the field of natural language processing, it is less conducive to tasks where the goal is to compare specific quantities of ingredients. In order to at least partially ameliorate these difficulties, some of pre-processing of the data set was performed. Some data that was not pertinent to the task at hand was dropped entirely. Among these columns were the directions and the Named Entity Recognition (NER) data. Then, in order to get the ingredient data into a more usable form for analysis, recipe ingredients were transposed to become the columns of the new dataset, with the recipe names remaining as the first column. A best effort was given to get the quantities for each ingredient for each recipe where applicable, but due to the semi-structured data this task proved difficult, and some amount of manual labor for data sanitization was required.

Manual analysis then starts by cleaning the dataset's columns and entries. Characters like punctuation, backslashes, and asterisks are filtered from ingredient names. Recipe names and ingredient amounts have similar but less stringent filtering. Care was taken to ensure the cleaning did not change the meaning of entries. For example, a nice filter would be to lowercase all entries but this would obscure teaspoon (denoted with a lowercase t) versus tablespoon (denoted with a capital T). Additional filters include replacing zeros and empty entries with NaN as well as removing any columns with no entries after filtering.

Removing characters from ingredient names creates duplicate columns (e.g. the columns "tsp. flour" and "tsp flour" would both become "tsp flour"), which would create problems when merging columns later. To avoid this duplicate ingredients are appended with a monotonically incrementing counter (in the previous example there would be two columns, "tsp flour", "tsp flour.1", ... "tsp flour.69")

Additionally, column names that start with "plus", "+", "&" or "and" were removed. This is because columns with "and" like terms are assumed to be a part of an improperly parsed larger section of the recipe with another ingredient. For example, a column "plus 2 Tbsp flour" was most likely something along the lines of "1 tsp salt plus 2 Tbsp flour". Any entries in the "plus 2 Tbsp flour" are not actually flour and are some other unknown ingredient. Lastly, "the world's largest apple pie" recipe is dropped from the dataset due to its outrageous list of ingredients. The recipe calls for 227,359 teaspoons of sugar.

The bulk of the remaining analysis is done through a mix of automated and manual methods. Manually, CSV files containing each column and its estimated unit, quantity, and bulk multiplier were created. For example "cans of applesauce, cups, 1.5, 1" where the scalar value 1.5 can be used to put the unit "cans" into "cups". These inform the automated analysis of assumed values for the column. Entries that omit a unit or quantity will fall back on the column default value. Quantity and units are converted to a common unit (teaspoon) before being inserted into the final dataset.

The final prepared dataset is then sorted into categories based on the recipe title. Four groups are created. American Apple Pies and UnAmerican Apple Pies, in addition to Dutch Apple Pies and unDutch Apple Pies as a control group.

Results and Conclusion

The results of the analysis are presented for the Dutch control group first in Table 1 and are plotted in Figure 1 as a stacked bar chart to display the percentage contribution of each key ingredient.

Table 1:Dutchness

Dutchness of Apple Pie (units in tsp)											
Apple Pie	Allspice	Apples	Butter	Cinnamon	Cloves	Dairy	Flour	Nutmeg	Salt	Shortening	Sugar
Dutch	0.014	159.270	27.261	1.089	0.006	11.899	86.064	0.592	0.482	1.999	125.366
UnDutch	0.060	147.649	29.229	1.211	0.090	24.809	64.458	0.506	0.650	10.035	94.849

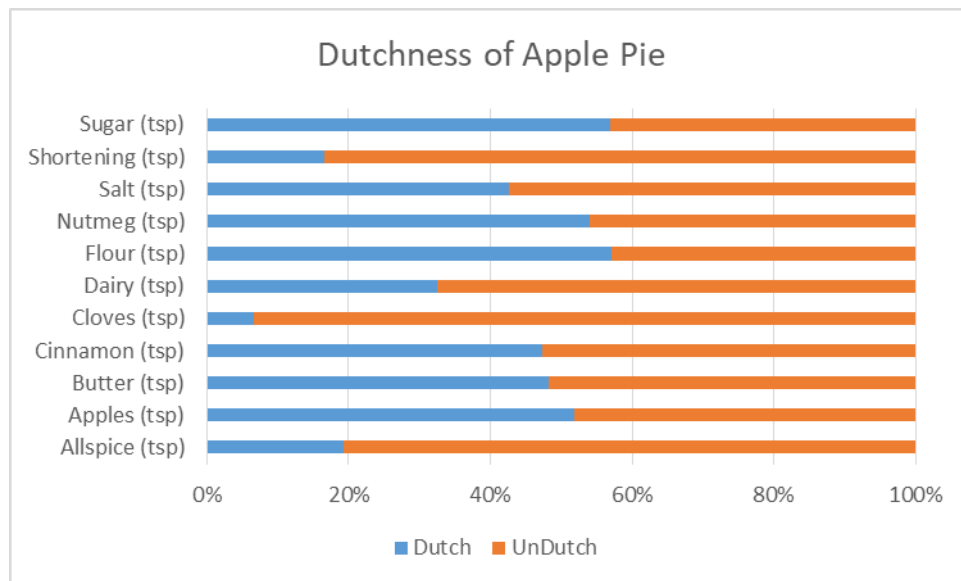


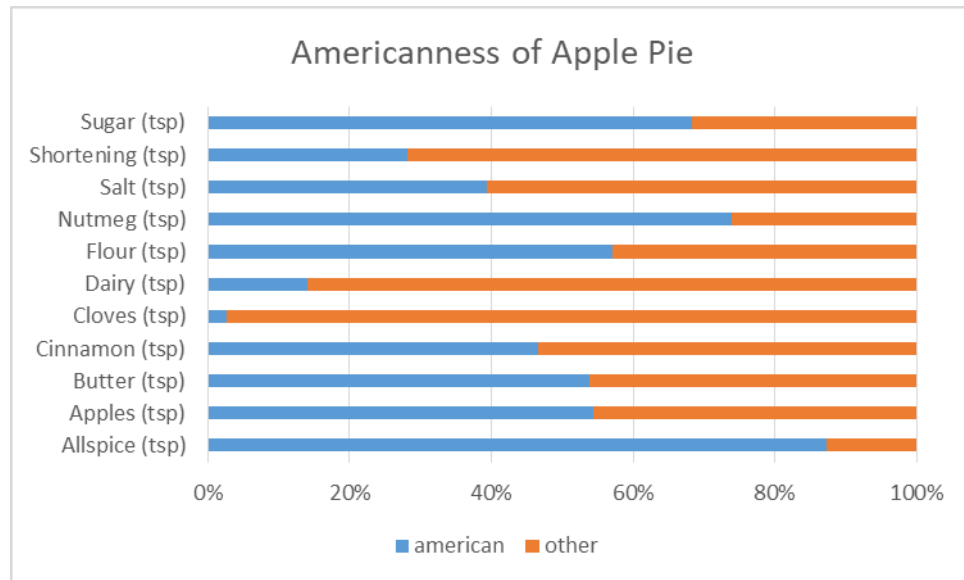
Figure 1: Dutchness Plotted

It can be seen that Dutchness is not characteristic of an excess of any one ingredient, as the Dutch apple pie has average or less than average amounts of each key ingredient. The main traits of Dutchness are a minimal amount of shortening, allspice, and cloves. The Dutch Apple Pie is included as a control group namely because of its signature construction, being invariably topped with a crumble made from butter, flour, and sugar. This crumble would presumably show up in the data when compared to UnDutch Apple Pies, and indeed it does. The Dutch Apple Pie has an above average amount of sugar and flour and an average amount of butter, which may perhaps be accounted for by the more austere Dutch taste

Having validated the pie measurement methodology we proceed to the important question at hand, the determination of Americanness. The results are listed in Table 2 and plotted in Figure 2.

Table 2: Americanness

Americanness of Apple Pie (units in tsp)											
Apple Pie	Allspice	Apples	Butter	Cinnamon	Cloves	Dairy	Flour	Nutmeg	Salt	Shortening	Sugar
American	0.382	175.750	33.913	1.059	0.002	4.000	86.106	1.422	0.419	3.885	204.059
UnAmerican	0.056	147.707	29.141	1.209	0.088	24.659	64.817	0.500	0.647	9.885	94.708

**Figure 2: Americanness Plotted**

Perhaps unsurprisingly one of the key characteristics of Americanness is an excess of sugar. American pies also tend to use an average amount of butter and a lesser amount of shortening. They tend to be spiced heavily with a gratuitous amount of allspice and nutmeg and yet contradictorily completely devoid of cloves. Most surprisingly, American apple pies use little dairy compared to UnAmerican pies, as the American Midwest which is most known for its pies is also a top dairy producer and consumer.

Not all of the sorted recipes contained every key ingredient and the number of recipes for each category are listed in Table 3 below, in the hopes that it may assist in future research. The total number of recipes was used when calculating the average quantity of ingredients in the respective categories.

Table 3: Quantity of Recipes

Number of Recipes in Each Category that Contain the Specified Ingredient												
Apple Pie	Allspice	Apples	Butter	Cinnamon	Cloves	Dairy	Flour	Nutmeg	Salt	Shortening	Sugar	TOTAL
American	8	49	39	42	1	12	35	31	20	8	43	52
UnAmerican	234	5108	4176	4180	204	1589	3805	1902	2088	537	4967	6319
Dutch	7	127	123	112	4	38	118	43	54	11	135	157
UnDutch	235	5030	4092	4110	201	1563	3722	1890	2054	534	4875	6214

Conclusion

The Apple Pie is as much a venerated symbol of America as the Statue of Liberty or the much celebrated “hot dog”. But the pie, unlike these other renowned symbols has a distinct breed of Americanness. Passed down from the depths of antiquity, the apple and the pie have traveled and evolved with humanity itself through the ages. Until now, where in our modern age the American apple pie has become a symbol festooned with the very traditions, history, and essence of America today.

As our great country weathers struggles and storms, we harken back to the Apple Pie to remind us of who we are, and what binds us all together in Americanness. And that ingredient, fellow Americans is sugar.

Coming in at a whopping 204 teaspoons (4 ¼ cups), the American apple pie uses more than twice as much sugar as the UnAmerican Pie. It also uses a significant amount of spice with its sugar, with allspice and nutmeg significantly exceeding the average and evoking flavors of the celebrated pumpkin pie as well. But perhaps with the recent divisions in America, what can most draw us together as a nation is our shared hatred for cloves.

Recipe

In order to taste the distilled Americanness as concluded above, the Authors fed the paper into Chat-GPT and asked it to create the most American Apple Pie Recipe using the ingredients in table 2. That recipe follows:

****The Most American Apple Pie Recipe****

Ingredients:

- Apples: 175.75 teaspoons (which is about 5.84 cups or approximately 6 medium-sized apples, peeled, cored, and sliced)
- Sugar: 204.059 teaspoons (which is about 4.25 cups, to sweeten your taste of freedom)
- Allspice: 0.382 teaspoon (just a dash for that homegrown kick)
- Nutmeg: 1.422 teaspoons (for that warm, spicy American hug)
- Cinnamon: We'll use 2 teaspoons here, because Americans love their cinnamon
- Butter: 33.913 teaspoons (which is about 7.1 ounces or a little less than 1 cup, for that rich, flaky crust)
- Flour: 86.106 teaspoons (which is about 1.8 cups, for the crust)
- Salt: a pinch (let's say 1/2 teaspoon, for the crust)
- Water: enough to make the dough workable (let's use 4 to 6 tablespoons, depending on the dough consistency)

Instructions:

1. **Prepare the Pie Crust:**

- In a large mixing bowl, whisk together about 1.8 cups of flour with 1/2 teaspoon of salt.
- Cut in the butter (just under 1 cup) until the mixture resembles coarse crumbs.
- Sprinkle in ice water, 1 tablespoon at a time, mixing until the dough comes together in a ball.

2. ****Make the Filling:****

- In a large bowl, combine about 6 cups of sliced apples with 4.25 cups of sugar, 0.382 teaspoon of allspice, 1.422 teaspoons of nutmeg, and 2 teaspoons of cinnamon. Stir them up until well mixed.

3. ****Assemble the Pie:****

- Roll out half of the dough on a floured surface to fit your pie plate.
- Transfer the dough to the plate, and trim the edges.
- Pour in the apple filling, heaping it in the center.
- Roll out the remaining dough and place it over the filling. Trim the overhang, and crimp the edges to seal.

4. ****Bake the Pie:****

- Preheat your oven to 425°F (220°C).
- Bake the pie for 15 minutes, then reduce the oven temperature to 350°F (175°C) and continue baking for another 45 minutes, or until the crust is golden brown and the filling is bubbly.

5. ****Cool and Serve:****

- Let the pie cool on a rack before slicing. This gives the filling time to set.

Pie

The Authors attempted to follow the recipe and bake the most American apple pie, Figure 3. However, fully embracing the spirit of freedom, they chose to not use the full 4 cups of sugar. Nevertheless, it was delicious.



Figure 3: American Apple Pie

Future work

The authors acknowledge that certain parsing errors and the small sample size for the American Apple Pie dataset may have improperly weighted the results. Certain errors in the base dataset were difficult to control for, like “34” being used instead of “3/4” which could be a key factor in why some ingredient amounts seem rather large. Further study is necessary to confirm the preliminary results presented here.

Additionally, with the computed dataset, there are many opportunities for further pie-related research. Particularly of interest is determining the qualities that define the “Best” apple pie, the “Original” apple pie, and most importantly, “Grandma’s” apple pie.

The analysis code and files are available: <https://github.com/YurBoiRene/apple-pie>

Works Cited

- Amandine Cornille, P. G.-R.-G. (2012). New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties. *PLOS Genetics*.
- Bien', M., Gilski, M., Maciejewska, M., Taisner, W., Wisniewski, D., & Ławrynowicz, A. (2020). RecipeNLG: A Cooking Recipes Dataset for Semi-Structured Text Generation. *Proceedings of The 13th International Conference on Natural Language Generation* (pp. 22-28). Dublin, Ireland: Association for Computational Linguistics.
- Bunschoten, R. C. (2014). *As American as Apple Pie: The History of American Apple Pie and Its Development into a National Symbol*. Bard College.
- Cao, Y., Kementchedjheva, Y., Cui, R., Karamolegkou, A., Zhou, L., Dare, M., . . . Hershcovich, D. (2024). Cultural Adaptation of Recipes. *Transactions of the Association for Computational Linguistics*, 80-99.
- Fatemi, B., Duval, Q., Girdhar, R., Drozdal, M., & Romero-Soriano, A. (2023). *Learning to Substitute Ingredients in Recipes*.
- Folger, J., & Thomson, S. (1923). *The Commercial Apple Industry of North America*. New York: The Macmillian Company.
- God. (1400 BC). *The Bible*. Jerusalem: Holy Catholic Church.
- Marín, J., Arito, B., Ferda, O., Nicholas, H., Amaia, S., Yusuf, A., . . . Antonia, T. (2019). Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*.
- Salvador, A., Drozdal, M., Giro-i-Nieto, X., & Romero, A. (2019). Inverse Cooking: Recipe Generation From Food Images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA: IEEE.
- Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE.
- Simmons, A. (1796). *American Cookery*.