JFC-Recipe: A Dataset for Nutrient Estimation from Japanese User-Generated Cooking Recipes

Keisuke Shirai[1,*], **Yoko Yamakata**[2,*,†], **Hirotaka Kameko**[1,*],
Akiko Sunto[3], **Jun Harashima**[4], **Shinsuke Mori**[1]
[1]Kyoto University, Kyoto, Japan [2]The University of Tokyo, Tokyo, Japan
[3]Kanagawa University of Human Services, Kanagawa, Japan
[4]Independent Researcher, Aomori, Japan
shirai.keisuke.l82@kyoto-u.jp, yamakata@hal.t.u-tokyo.ac.jp,
{kameko, forest}@i.kyoto-u.ac.jp, suntou-a@kuhs.ac.jp, j.harashima@gmail.com

Abstract

Estimating nutrients from recipes is essential for performing proper daily dietary control. The nutrients of the recipe could be roughly calculated by identifying the nutrients and weights of each ingredient in the recipe. However, no dataset with fully manual annotations of nutritional values and weights has been released so far, especially for Japanese recipes. In this work, we propose a novel dataset called the Japanese Food Composition Recipe Dataset (JFC-Recipe). The JFC-Recipe dataset consists of two types of annotations: (i) food item annotation that links ingredients in recipes to a database providing nutrients for foods and (ii) amount and unit annotation that are converted into weights in grams using a weight table. We describe a data collection procedure and annotation process, show statistics, and provide inter-annotator agreements to validate the quality of our annotations. In experiments, we tackle two tasks of food item estimation and quantity estimation. Experimental results show that pre-trained language models learn to estimate food items and quantities accurately.

Keywords: annotation, cooking recipe, nutrient estimation

1. Introduction

Daily dietary habits have a significant impact on disease risk and life expectancy. In recent years, there has been a growing interest in building a system that provides a healthy diet (Afshin et al., 2019; Willett et al., 2019). Since dishes are inherently based on cooking recipes, estimating nutrients from recipes is a promising way to build a dietary control system. Compared to recipes created by professionals, user-generated recipes exist in larger quantities and are more closely tied to our daily lives as they are easily accessible through recipe websites and social media. Therefore, this study focuses on nutrient estimation from user-generated recipes.

In natural language processing, there exist few studies on estimating nutrients from recipes. Existing work (Harashima et al., 2020; Feher et al., 2023) built a private database that consists of normalized food expressions and nutrients and proposed to normalize ingredient expressions to link them to the database. However, these approaches have issues with the comprehensiveness of the foods and the maintenance and management costs associated with the database. On the other hand, there exist databases that are developed and maintained by national authorities. These databases can be considered more reliable in terms of comprehensiveness, quality, and future extensibility. Therefore, the above issues can be remedied with such databases.

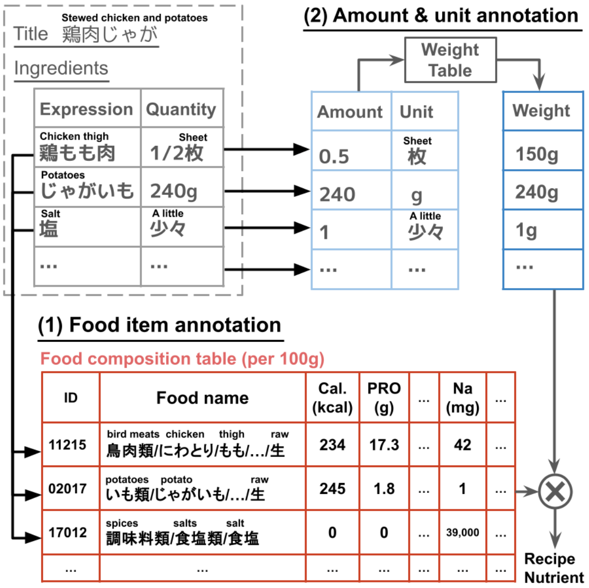

Figure 1: Overview of our annotation. (1) Ingredient to food name linking annotates ingredient expressions in recipes with food items in the food composition database. (2) Quantity annotation annotates the ingredients with amounts and units, which are converted into weights automatically using a weight table.

---

*Equal contribution
† Corresponding author

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 6349–6360
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

Food	The Japanese food composition table	The USDA database
skin-on chicken thigh	$\overset{\text{bird meats}}{\text{鳥肉類}}$/$\overset{\text{chicken}}{\text{にわとり}}$/$\overset{\text{thigh}}{\text{もも}}$/$\overset{\text{skin}}{\text{皮つき}}$/$\overset{\text{raw}}{\text{生}}$	Chicken, breast, meat and skin, raw
skinless potato	$\overset{\text{potatoes}}{\text{いも類}}$/$\overset{\text{potato}}{\text{じゃがいも}}$/$\overset{\text{tuber}}{\text{塊茎}}$/$\overset{\text{without skin}}{\text{皮なし}}$/$\overset{\text{raw}}{\text{生}}$	Potatoes, gold, without skin, raw

Table 1: Examples of food items in the Japanese food composition table and the USDA database. Food items are defined at a similar granularity in terms of food type and state.

In addition to linking ingredient expressions to the database, quantity estimation for ingredients (Choi et al., 2023) needs to be addressed. Especially in user-generated recipes, these expressions are diverge due to inconsistencies in notation and typographical errors. For example, the amount "1/2" may be written as "0.5" or "½." Units may sometimes be omitted, but actual weight can vary greatly depending on the unit (e.g., "1 cup" and "1 teaspoon"), so it should be estimated accurately. Therefore, estimating the amount and unit is simultaneously important with linking ingredients to the database.

To tackle the above challenges, we propose Japanese Food Composition Recipe (**JFC-Recipe**), a novel dataset for nutrient estimation from recipes. As shown in Figure 1, the JFC-Recipe dataset consists of two types of annotations for ingredients: (i) food item annotation and (ii) amount and unit annotation. Our annotation is based on $1,985$ Japanese recipes from the Cookpad dataset (Harashima et al., 2016). For the food item annotation, we adopt the Standards Tables of Food Composition in Japan, maintained by the Japanese government, and we refer to it as the food composition table in this study. The food composition table consists of food items that provide nutrients (e.g., cal., PRO, FAT, CHO) per 100g for ingredients with a hierarchical food name (e.g., "$\overset{\text{mammalian meats}}{\text{畜肉類}}$/$\overset{\text{beef}}{\text{うし}}$/$\overset{\text{Japanese beef cattle}}{\text{和牛肉}}$/$\overset{\text{rib loin}}{\text{リブロース}}\overset{\text{lean}}{\text{赤肉}}$/$\overset{\text{raw}}{\text{生}}$"). We manually annotated all ingredients in the recipes with food items. For the quantity annotation, we annotated the ingredients with an amount (e.g., "2") and unit (e.g., "$\overset{\text{stick}}{\text{本}}$") separately. We also built a weight table that enables us to convert amounts and units into weights in grams. To our knowledge, this is the first Japanese recipe dataset that provides both ingredient nutrients and quantities that consist of amounts and units.

We tackle nutrient estimation from recipes by dividing the task into two sub tasks: (i) food item estimation and (ii) quantity estimation. We formulate these tasks and investigate estimation performance by pre-trained language model-based methods. Experimental results show that our methods achieve higher performance as the training data size increases, demonstrating the scalability of our framework.

Our contributions are summarized as:

- We propose the JFC-Recipe dataset that consists of (i) food item annotation and (ii) amount and unit annotation for ingredients in recipes. We first describe a data collection criterion and annotation process, and then provide statistics and inter-annotator agreements.
- We formulate two tasks of food item estimation and quantity estimation for nutrient estimation from recipes. We provide experimental results using pre-trained language model-based methods.

2. Related Work

A cooking recipe is a structured text and has been used for a wide range of research in natural language processing. Procedures in a recipe have been represented as a graph structure (Mori et al., 2014; Kiddon et al., 2015; Jermsurawong and Habash, 2015; Yamakata et al., 2020), and methods to automatically estimate the graphs have also been developed (Maeta et al., 2015; Kiddon et al., 2015; Jermsurawong and Habash, 2015; Bhatt et al., 2024). Executing steps causes ingredient state changes, and estimating the changes has been addressed to consider causal effects of actions on entities (Bosselut et al., 2018; Tandon et al., 2020; Kim and Schuster, 2023). Recipe translation between languages (Sato et al., 2016) and cultures (Cao et al., 2024) has also been conducted. Furthermore, there exist studies considering multimodality, such as step generation from visual information (Ushiku et al., 2017; Zhou et al., 2018; Salvador et al., 2019; Nishimura et al., 2019), cross-modal learning between recipes and dish images (Salvador et al., 2017; Chen et al., 2018; Carvalho et al., 2018; Salvador et al., 2021), and dataset creation for multimodal recipe understanding (Hashimoto et al., 2014; Nishimura et al., 2020; Pan et al., 2020; Shirai et al., 2022; Maeda et al., 2024).

While cooking recipes have been used for a wide variety of studies, estimating nutrients from recipes has not been extensively studied, compared to approaches that estimate nutrients from dish images (Meyers et al., 2015; Thames et al., 2021). Estimation from recipes has advantages such as (i) knowing nutrients of dishes before cooking and (ii) handling ingredients that are visually challenging to recognize, such as seasonings. Inspired by previous work, we tackle nutrient es-

ID	Food name	Cal. (kcal)	PRO (g)	FAT (g)	CHO (g)	Na (mg)	Ca (mg)	···
1031	$\overset{\text{wheat}}{\text{こむぎ}}$/$\overset{\text{breads}}{\text{パン類}}$/$\overset{\text{French bread}}{\text{フランスパン}}$	289	9.4	1.3	57.5	620	16	···
11008	$\overset{\text{mammalian meats}}{\text{畜肉類}}$/$\overset{\text{beef}}{\text{うし}}$/$\overset{\text{dairy fattened steer}}{\text{乳用肥育牛肉}}$/$\overset{\text{chuck roll}}{\text{かたロース}}$/$\overset{\text{lean and fat}}{\text{脂身つき}}$/$\overset{\text{raw}}{\text{生}}$	295	16.2	26.4	0.2	50	4	···
11013	$\overset{\text{mammalian meats}}{\text{畜肉類}}$/$\overset{\text{beef}}{\text{うし}}$/$\overset{\text{Japanese beef cattle}}{\text{和牛肉}}$/$\overset{\text{rib loin}}{\text{リブロース}}$/$\overset{\text{lean}}{\text{赤肉}}$/$\overset{\text{raw}}{\text{生}}$	395	14.0	40.0	0.2	53	3	···

Table 2: Examples of food items in the food composition table. Calories, protein (PRO), fat (FAT), carbohydrate (CHO), minerals (e.g., Na), and vitamins (e.g., Ca) show nutritional values per 100g.

timation from recipes in combination of linking ingredients to the database (Harashima et al., 2020; Feher et al., 2023) and estimating amounts and units for ingredients (Choi et al., 2023).

Recipe1M+ (Marın et al., 2021) is a dataset having similar annotations to our dataset, but there are the following three major differences between Recipe1M+ and our JFC-Recipe dataset. (1) A portion of Recipe1M+ has ingredients annotated with food items in a database[1] maintained by the United States Department of Agriculture (USDA). As shown in Table 1, this database defines food items with a granularity comparable to that of the Japanese food composition table used in this study. Ingredients in Recipe1M+ are assigned food items by text matching, suffering from selection bias. This bias is avoided in our dataset by manually annotating all ingredients. This claim is supported by the number of unique food items: Our annotation has 666 types, while Recipe1M+ has only 357. (2) Since quantity expressions in Recipe1M+ are extracted by pattern matching, targetable quantity expressions are limited. We manually annotated amounts and units and created a weight table to convert them into weights in grams. (3) The JFC-Recipe dataset is the first Japanese dataset for nutrient estimation from recipes. Since food culture varies by culture, building datasets across diverse cultures is valuable.

Choi et al. (2023) proposed KitchenScale, a model that estimates the quantities of ingredients in recipes. The objective of the task is to estimate the target ingredient's measurement type (volume or weight), unit, and amount. The task is based on three subtasks: measurement type classification, unit classification, and amount regression. Our proposed nutrient estimation approach is similar to KitchenScale in that it estimates the units and amounts of ingredients. Our objective is to estimate the nutrients of recipes, and since the nutrients of each food item are provided per 100g, it is necessary to estimate the weights of ingredients. Therefore we propose to create a weight table that provides weights for each food item per unit and estimates the weights of ingredients by using the weight table.

3. Preliminaries

This section provides the definitions of terms that we use throughout this paper and describes preliminary knowledge.

3.1. Cooking Recipe

We consider that a cooking recipe consists of the following components: (i) a recipe title t, (ii) a list of M ingredients $r_1, r_2, \cdots, r_M$, (iii) quantities for the ingredients $q_1, q_2, \cdots, q_M$, and (iv) a number of servings.[2] Each ingredient r_i (e.g., “$\overset{\text{cucumber}}{\text{きゅうり}}$”) is paired with a quantity q_i (e.g., “１$\overset{\text{stick}}{\text{本}}$”). A quantity generally consists of an amount (e.g., “１”) and a unit (e.g., “$\overset{\text{stick}}{\text{本}}$”), where units may be omitted. All the components are written in natural language.

3.2. Food Composition Table

There exist databases that are maintained by national authorities and define nutrients for food ingredients, such as the Standards Tables of Food Composition in Japan (Japan), FoodData Central (U.S.), and food composition tables for China (China). Since this study targets Japanese recipes, we use the latest version of the Standards Tables of Food Composition in Japan[3], which is maintained by the Japanese government and defines 2,538 types of foods. Throughout this paper, we refer to the table as the food composition table and each record as a food item. As shown in Table 2, the food composition table describes the nutrients contained per 100g. Each food item can be uniquely identified by either its ID or its food name, which is a hierarchical notation for foods.

4. The JFC-Recipe Dataset

The JFC-Recipe dataset is characterized by two types of annotations for ingredients in recipes: (i) food item annotation and (ii) amount and quantity

---

[1] Standard Reference, release 27, containing 8,618 food items.

[2] Other components, such as steps and user review, are not considered in this study.

[3] `https://www.mext.go.jp/a_menu/syokuhinseibun` (Japanese). Accessed (for all URLs): 2025-10-16.

Rank	Expression	Frequency	Coverage	ID	Food name
1	$\overset{\text{sugar}}{\text{砂糖}}$	427,440	3.359%	3003	$\overset{\text{sugars}}{\text{砂糖類}}$ / $\overset{\text{soft sugar}}{\text{車糖}}$ / $\overset{\text{white soft sugar}}{\text{上白糖}}$
2	$\overset{\text{salt}}{\text{塩}}$	342,152	2.768%	17012	$\overset{\text{seasonings}}{\text{調味料類}}$ / $\overset{\text{edible salts}}{\text{食塩類}}$ / $\overset{\text{salt}}{\text{食塩}}$
3	$\overset{\text{Water}}{\text{水}}$	307,591	2.417%	N/A	$\overset{\text{water}}{\text{水}}$
4	$\overset{\text{soy sauce}}{\text{醤油}}$	250,369	1.966%	17007	$\overset{\text{seasonings}}{\text{調味料類}}$ / $\overset{\text{soy sauces}}{\text{しょうゆ類}}$ / $\overset{\text{dark soy sauce}}{\text{こいくちしょうゆ}}$
5	$\overset{\text{egg}}{\text{卵}}$	243,772	1.916%	12004	$\overset{\text{chicken egg}}{\text{鶏卵}}$ / $\overset{\text{whole egg}}{\text{全卵}}$ / $\overset{\text{raw}}{\text{生}}$
6	$\overset{\text{sake}}{\text{酒}}$	197,019	1.548%	17138	$\overset{\text{seasonings}}{\text{調味料類}}$ / $\overset{\text{others}}{\text{その他}}$ / $\overset{\text{cooking sake}}{\text{料理酒}}$
7	$\overset{\text{milk}}{\text{牛乳}}$	173,950	1.367%	13003	$\overset{\text{milks}}{\text{牛乳及び乳製品}}$ / $\overset{\text{liquid milks}}{\text{液状乳類}}$ / $\overset{\text{regular milk}}{\text{普通牛乳}}$
8	$\overset{\text{mirin}}{\text{みりん}}$	153,162	1.204%	16025	$\overset{\text{alcohols}}{\text{アルコール飲料類}}$ / $\overset{\text{liqueurs}}{\text{混成酒類}}$ / $\overset{\text{mirin}}{\text{みりん}}$ / $\overset{\text{hon mirin}}{\text{本みりん}}$
9	$\overset{\text{butter}}{\text{バター}}$	152,100	1.195%	14017	$\overset{\text{butters}}{\text{バター類}}$ / $\overset{\text{unfermented butter}}{\text{無発酵バター}}$ / $\overset{\text{salted butter}}{\text{有塩バター}}$
10	$\overset{\text{onion}}{\text{玉ねぎ}}$	148,513	1.167%	6153	$\overset{\text{onions}}{\text{たまねぎ類}}$ / $\overset{\text{onion}}{\text{たまねぎ}}$ / $\overset{\text{bulb}}{\text{りん茎}}$ / $\overset{\text{raw}}{\text{生}}$

Table 3: Top 10 ingredient expressions found in the Cookpad dataset. The ingredient expressions are annotated with food items. The ID for “$\overset{\text{water}}{\text{水}}$” is annotated with N/A (Not Available) because it is not defined in the food composition table.

annotation. This section is structured as follows: a data collection procedure (Section 4.1), a food item annotation (Section 4.2), amount and unit annotation (Section 4.3), statistics (Section 4.4), and inter-annotator agreement (Section 4.5).

4.1. Data Collection

We selected recipes from the Cookpad dataset (Harashima et al., 2016), which provides approximately 1.7 million user-generated recipes uploaded on Cookpad[4] and is available for use in academic research. Since Cookpad is the largest service for providing Japanese recipes, the Cookpad dataset is expected to cover a wide variety of recipes. Recipes in the Cookpad dataset consist of a recipe title, ingredient list, step-by-step cooking instructions, and other metadata (e.g., number of servings). For annotation, we used 2,000 recipes from the first two sections of the Cookpad dataset.

4.2. Food Item Annotation

As described in Section 3.2, the food composition table provides food items that defines nutrients per 100g for foods. We annotated ingredients in the 2,000 recipes (Section 4.1) with food items. For efficient annotation, we first built an automatic estimator to estimate frequently appearing food items and then instructed annotators to check and correct the annotated results.

4.2.1. Frequent Food Item Selection

We first extracted ingredient expressions in all recipes in the Cookpad dataset. This resulted in 959,543 types of unique expressions.[5] We selected the top 2,000 expressions, and asked nutritionists to annotate them with corresponding food items. In cases where no relevant food items existed, we requested that N/A (Not Available) be assigned. As a result, 587 types of food items were annotated. Table 3 shows the top 10 ingredient expressions annotated with food items.

4.2.2. Annotation

Based on the 2,000 expressions with the 587 food names, we built a food item estimator that classifies given ingredient expressions into one of 587 classes by training a pre-trained Japanese BERT[6]. Ingredient expressions in the collected 2,000 recipes were automatically annotated with food items estimated by the estimator. We then instructed annotators to manually check the estimated food items and correct if the items were inappropriate for the recipe. During this annotation, the annotators were asked to use all $2,538$ food items and two additional labels of “water” and “N/A”. We finally used 668 labels, which include the two additional labels. During annotation, we excluded recipes with poorly written ingredients and those that were not actual cooking recipes. As a result, 1,985 recipes were annotated.

4.3. Amount and Unit Annotation

We first annotated the ingredients in the 1,985 recipes with amounts and units. We then created

---

[4] `https://cookpad.com/` (Japanese).

[5] We found that 773,188 (80.6%) expressions only appear once, showing the diversity of ingredient expressions in user-generated recipes.

[6] `https://huggingface.co/tohoku-nlp/bert-base-japanese-whole-word-masking`

Food item		Gram per unit category									
ID	Food name	no unit information 無単位	sheet 枚	stick 本	···	whole 個片丁株玉房	cup 杯/カップ	half 半分	···	handful 一掴	bag 袋	···
3003	sugars refined sugar caster sugar 砂糖類/車糖/上白糖	3	0	0	···	0	120	0	···	25	1,000	···
6065	cucumber fruit raw きゅうり/果実/生	100	0	100	···	100	0	50	···	0	0	···
7107	banana raw バナナ/生	0	0	120	···	120	0	0	···	0	0	···
11215	bird meats chicken thigh skin raw 鳥肉類/にわとり/もも/皮付き/生	0	280	0	···	28	0	0	···	0	0	···
12005	chicken egg whole egg raw 鶏卵/全卵/生	60	0	0	···	30	0	30	···	0	0	···

Table 4: Examples from the created weight table. Cells represent weights for the combination of food items and units. We set 0 for combinations that did not appear during annotation.

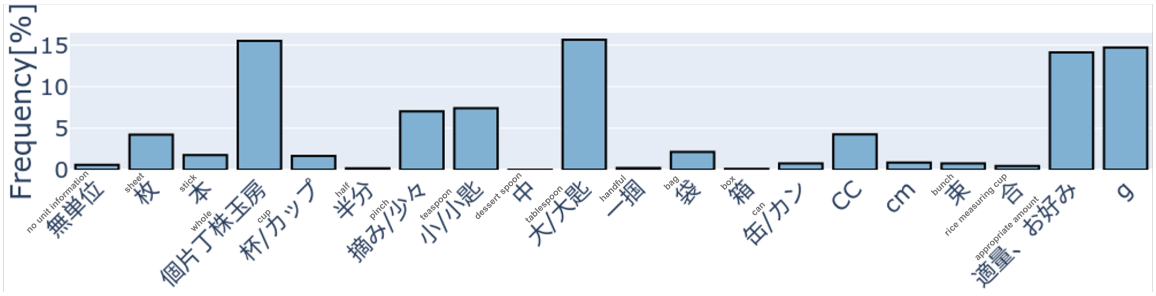

Figure 2: Frequency of annotated units.

a weight table that enables us to convert amounts and units into weights in grams.

4.3.1. Annotation

To maintain the consistency of units, we defined the following 20 categories for units: “無単位” (no unit information), “枚” (sheet), “本” (stick), “個片丁株玉房” (whole), “杯/カップ” (cup), “半分” (half), “摘み/少々” (pinch), “小/小匙” (teaspoon), “中” (dessert spoon), “大/大匙” (tablespoon), “一掴” (handful), “袋” (bag), “箱” (box), “缶/カン” (can), “CC”, “cm”, “束” (bunch), “合” (rice measuring cup), “適量、お好み” (appropriate amount), and “g” (gram). Note that units with the same or similar meaning are grouped into a single category. We instructed annotators[7] to annotate the ingredients with a pair of amount and unit category while referring to original quantity expressions in the recipes.

4.3.2. Weight Table Creation

We manually created a weight table that consists of combinations of all 2,538 food items and the 20 units defined in Section 4.3.1. Each combination represents weights in grams. The annotators were instructed to search the web for the weight corresponding to each combination and assign an average value. The combinations that did not appear during annotation were assigned 0. Table 4 shows examples. In sugar (ID 3003), for example, “杯/カップ” (cup) is assigned 120g, “一掴” (handful) is assigned 25g, and “袋” (bag) is assigned 1,000g.

Rank	Food item	Count
1	seasonings edible salts salt 調味料類/食塩類/食塩	822
2	N/A	794
3	sugars soft sugar caster sugar 砂糖類/車糖/上白糖	572
4	seasonings soy sauces dark soy sauce 調味料類/しょうゆ類/こいくちしょうゆ	540
5	water 水	454
6	spices pepper black powder 香辛料類/こしょう/黒/粉	393
7	chicken egg whole egg raw 鶏卵/全卵/生	384
8	onions onion bulb raw たまねぎ類/たまねぎ/りん茎/生	321
9	seasonings others cooking sake 調味料類/その他/料理酒	311
10	wheat wheat flour cake flour first grade こむぎ/小麦粉/薄力粉/1等	261

Table 5: Top 10 food items annotated in the JFC-Recipe dataset.

4.4. Statistics

The annotated 1,985 recipes contained 15,188 ingredients, all of which were annotated with food items, amounts, and units. Table 5 shows the top 10 food items in our annotation. The appearing food items are similar to those in Table 3, confirming that the recipes were randomly sampled from the Cookpad dataset. We also found that the number of unique food items per recipe was smaller than that of ingredients per recipe. This is because some ingredients in a single recipe may be annotated with the same food item (e.g., “salt” may appear multiple times in the ingredient list for different purposes). One might be concerned that the dataset includes ingredients with no nutritional values. In our annotation, only 水 (water) and N/A are de-

---

[7] We asked Japanese annotators who cook on a daily basis and can prepare food by reading web recipes.

Annotation type	Agreement
Food item	75.5%
Amount	89.0%
Unit	88.2%

Table 6: Inter-annotator agreements. Cohen's Kappa is used for the food item and unit, while the agreement for the amount is calculated based on the difference in numerical values.

fined to provide no nutrients, and they account for 8.22% of the ingredients in our dataset (see Table 5). This indicates that the remaining 92% of food items provide nutritional values, including micronutrients, as shown in Table 2.

Figure 2 shows a histogram of annotated units. The frequency of each unit is not equally distributed, with four units accounting for 60% of the total: “$\overset{\text{whole}}{\text{個片丁株玉房}}$”, “$\overset{\text{tablespoon}}{\text{大/大匙}}$”, “$\overset{\text{appropriate amount}}{\text{適量、お好み}}$”, and “g”. Interestingly, “g” accounts for only 15% of the total units, and this highlights the importance of estimating weights from units.

4.5. Inter-annotator Agreement

To investigate the quality of the annotations, we sampled 100 recipes (5%) from the dataset, instructed other annotators to annotate the samples, and then calculated the inter-annotator agreements. This annotation consists of the three annotation types: food items, amounts, and units. For measurements, we used Cohen's Kappa (Cohen, 1960) for the food items and units, which handle categorical variables. For the amounts, we calculated the agreement rate by considering that two annotated amounts were the same if both were labeled as N/A or the difference of the numerical values was 0.01 or less.

Table 6 shows the agreements. The agreements for the food items and units were 75.5% and 88.2%, respectively, and these values are excellent by Fleiss' guideline (Fleiss et al., 2013). For the amounts, it achieved a high agreement of 89.0%. These agreements indicate that our annotations are of high quality.

5. Experiments

With the JFC-Recipe dataset, we perform nutrient estimation from recipes. We divide the nutrient estimation into **food item estimation** (Section 5.1) and **quantity estimation** (Section 5.2) and investigate estimation performance in these tasks using pre-trained language models.

5.1. Food Item Estimation

Food item estimation aims to estimate corresponding food items to ingredient expressions in a recipe. In this section, we formulate the task in Section 5.1.1, describe methods in Section 5.1.2, explain experimental settings in Section 5.1.3, and provide results in Section 5.1.4.

5.1.1. Task Definition

Given a recipe with M food ingredient expressions $(r_1, r_2, \cdots, r_M)$, the goal of this task is to estimate corresponding food items $(t'_1, t'_2, \cdots, t'_M)$ to the ingredients. Food items must exist in a predefined set $\{t_1, t_2, \cdots, t_N\} (N > M)$, which also includes “water” and “N/A” (not applicable). As we described in Section 4.2.2, $N = 668$ in our experiments.

5.1.2. Estimation Models

Since target food items are in a closed set, fine-tuning a pre-trained encoder (e.g., BERT (Devlin et al., 2019)) for classification is a straightforward solution to this problem. On the other hand, fine-tuning a large language model (LLM) (e.g., LLAMA 3 (Grattafiori et al., 2024)) is now a de facto standard for solving NLP tasks. Thus, we consider two approaches of classification by BERT and generation by LLM.

Classification A pre-trained BERT first converts an ingredient expression $r^1 r^2 \cdots r^K$ into a sequence of encodings $h^1 h^2 \cdots h^K$. A mean pooling is performed to the encodings as $\hat{h} = \Sigma_{i=0}^{K} h^i$, and $\tilde{h} \in \mathbb{R}^N$ is calculated as: $\tilde{h} = W_2(\text{ReLU}(W_1 \hat{h}))$, where $W_1 \in \mathbb{R}^{d_2 \times d_1}$ and $W_2 \in \mathbb{R}^{N \times d_2}$ are linear transformations, and d_1 is the same size as the BERT model's hidden size. A class y is estimated as $y = \arg\max \tilde{h}$. We process all ingredients in a recipe simultaneously by concatenating the ingredient expressions as $\texttt{[CLS]}r_1\texttt{[INGR]}r_2\texttt{[INGR]}\cdots\texttt{[INGR]}r_M\texttt{[SEP]}$, where `[INGR]` is a newly added special token to separate ingredients. In our preliminary experiments, this boosts performance significantly compared to encoding ingredients independently.[8]

Generation A pre-trained LLM takes as input a sequence of ingredient expressions in a recipe and outputs a sequence of food items corresponding to the ingredients. We can represent inputs and outputs in data description language[9] to parse them easily. The model is trained by supervised fine-

---

[8] We consider that other ingredients in the recipe are used as additional contextual information, but we leave detailed analysis to future work.

[9] We adopted a JSON format in this work.

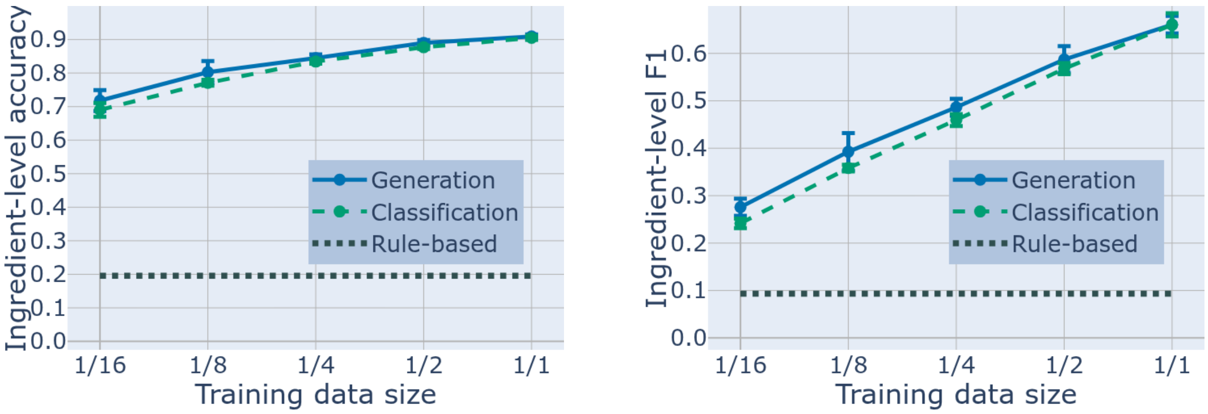

Figure 3: Ingredient-level accuracy (left) and macro F1 (right) by varying the training data size. Generation refers to a generative approach with a pre-trained LLM, Classification refers to a classification approach with a pre-trained BERT, and Rule-based is a rule-based method based on Levenshtein distance.

tuning to generate the outputs directly.[10] One concern with LLM is the possibility of generating nonexistent food items (Maynez et al., 2020) due to hallucination or confabulation. To address this, we use a constrained decoding strategy using a trie, inspired by previous work (Harashima and Yamada, 2018; Cao et al., 2021). A trie is built on tokenized food item expressions and is used to select valid tokens when generating food items. In our setting, a unique phrase (e.g., “food item:”) precedes food items, and thus it enables the model to automatically switch the constrained decoding on and off. Similarly to the classification model, we train the model to simultaneously estimate food items for all ingredients in a recipe.

5.1.3. Settings

Models Since our dataset consists of Japanese recipes, we adopted pre-trained models specialized for Japanese. Concretely, we used a character-level Japanese BERT model[11] for the classification model and an LLAMA 3 mid-trained on Japanese corpus (Fujii et al., 2024; Okazaki et al., 2024)[12].[13] We also demonstrate a rule-based baseline in addition to the LLM-based models. This baseline method selects a food name based on the Levenshtein distance between a food expression and food names in the food composition table.

Hyper-parameters and optimization For the classification model, we set $(d_1, d_2) = (1024, 1024)$. The all parameters were tuned by AdamW (Loshchilov and Hutter, 2019) with a constant learning rate of 1.0×10^{-5} and a minibatch size of 1. For the generation model, we used low-rank adaptation (Hu et al., 2022) to efficiently finetune the LLM. The parameters were tuned by AdamW with a minibatch size of 1. We tuned a learning rate by linearly increasing it from 0 to 5×10^{-4} for $1,000$ steps and then decreasing to 0 for $9,000$ steps with a cosine annealing. A weight decay factor was set to 0.01.

Metrics The dataset is split in a 7:1:2 ratio to create a training, validation, and test sets. To obtain reliable results, we created different splits for 5 times, train and evaluate models, and report the mean and standard deviation of scores. We used the following three metrics:

- **Ingredient-level accuracy** simply calculates the ratio of successfully estimated food items.
- **Ingredient-level F1** calculates the macro F1 of estimated food items. The use of F1 aims to account for minority classes.
- **Recipe-level accuracy** calculates the ratio of recipes in which all food items were successfully estimated. We consider this metric to evaluate the results at the recipe level.

5.1.4. Experimental Results

Figure 3 and Figure 4 show results by varying the training data size. The learning-based models significantly outperform the rule-based method even when only one-sixteenth of the training data is available. This result not only demonstrates

---

[10] Generating outputs in a wrong format might cause a parse error, but we confirmed that the model learns to generate in a valid format accurately (9̃9%) in our preliminary experiments.

[11] `https://huggingface.co/tohoku-nlp/bert-large-japanese-char-v2`

[12] `https://huggingface.co/tokyotech-llm/Llama-3-Swallow-8B-Instruct-v0.1`

[13] Performance of both models might be boosted by using a modernized architecture for BERT (Warner et al., 2025) and incorporating a retrieval augmented generation (Lewis et al., 2020) for the LLM, but we leave detailed analysis to future work.

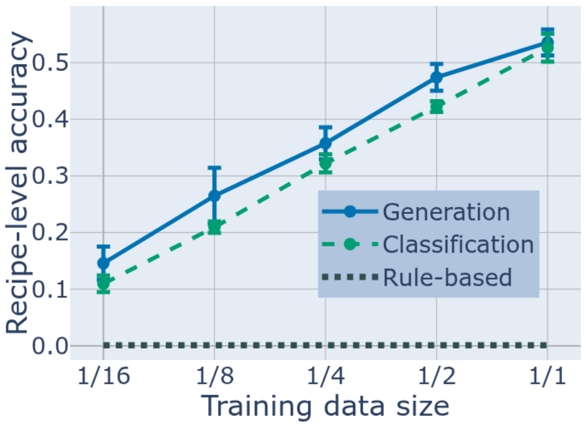

Figure 4: Recipe-level accuracy by varying the training data size.

the challenge of estimating food names from user-written ingredient expressions but also justifies the use of pre-trained language models. The learning-based models improve the scores as the training data size increases, indicating the scalability of this approach. Comparing both models, the generation model achieves better performance with smaller training data, but the classification model achieves competitive performance with full data. Since the inference speed of the classification model is faster than that of the generation model, this result implies that the classification model can be a practical choice when labeled data are abundant.

5.2. Quantity Estimation

In addition to the food items, weights for each food item are required to calculate the nutrients of the recipe. We obtain weights by estimating ingredient quantities that consist of amounts and units. Additionally, we also estimate the number of servings to calculate the nutrients per serving. Thus, we consider that the quantity estimation consists of two tasks: ingredient quantity estimation and servings estimation.

5.2.1. Task Definition

Given a recipe containing M food ingredient expressions $(x_1, x_2, \cdots, x_M)$, M user-written quantity expressions $(q_1^e, q_2^e, \cdots, q_M^e)$, M food items $(y_1, y_2, \cdots, y_M)$, and the serving expression s^e, the goal of this task is to estimate the quantities of each ingredient $(q_1, q_2, \cdots, q_M)$ and the number of servings s. Each quantity of ingredient q_i is expressed as a single number and a single unit (q_i^v, q_i^u), where q_i^v is a real number and q_i^u is one of the pre-defined units. The quantity is “N/A” (not applicable) when the quantity is not described or described as “an appropriate amount.”

The gold quantities q_i are from our annotation. The pseudo-gold number of servings s is extracted from the serving expression s^e by following rule-based methods: (1) extract all numbers in the expression, including Arabic numerals and Chinese numerals, (2) convert the numbers to Arabic numerals, and (3) use the average if multiple numbers are extracted. For example, “$\overset{\text{2 to 3 servings}}{\text{2〜3 人前}}$” is converted to 2.5.

5.2.2. Estimation Model

Similarly to the generation approach in Section 5.1, we fine-tune a pre-trained LLM to solve tasks in an end-to-end manner. Especially, we train the model to solve both ingredient quantity estimation and serving estimation using the same model. Ingredient quantity estimation is performed independently for each ingredient. The input prompts of each task are as follows:

Ingredient quantity estimation

Estimate the quantity of the following ingredient. Represent the quantity in a numerical value and a unit. The unit is in ’無単位’, ’枚’, ’本’, ’個片丁株玉房’, ’杯/カップ’, ’半分’, ’摘み/少々’, ’小/小匙’, ’中’, ’大/大匙’, ’一掴’, ’袋’, ’箱’, ’缶/カン’, ’CC’, ’cm’, ’束’, ’合’, ’適量、お好み’, or ’g’.

Serving estimation

Estimate the number of servings of the following recipe. Represent the number of servings in a numerical value.

5.2.3. Settings

Almost all settings are the same as those in section 5.1.3, except for the following points.

Models Since the task is not classification except unit estimation, we only evaluate the generation approach. We represent inputs and outputs in a TSV format to parse them easily.

Hyper-parameters and optimization The parameters were tuned with a minibatch size of 2.

Metrics We used the following two metrics:

- **Quantity estimation accuracy** calculates the ratio of ingredients in which both the amount and unit are correctly estimated.
- **Number of servings estimation accuracy** calculates the ratio of recipes in which the number of servings is correctly estimated.

5.2.4. Experimental Results

Figure 5 shows results by varying the training data size. In both tasks, the performance improves as

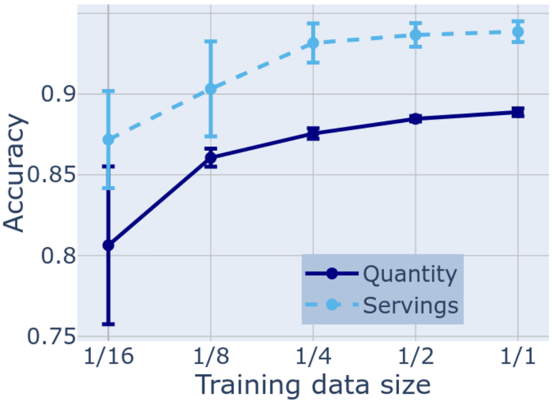

Figure 5: Accuracy of quantity estimation and number of servings estimation by varying the training data size.

the training data size increases, similarly to the results in Section 5.1.4. In the quantity estimation task, the model achieves 88.9% accuracy when using all training data (1/1), which is a reasonable result compared to the inter-annotator agreement (89.0% in amount estimation, 88.2% in unit estimation). In the number of servings estimation task, the model achieves 93.9% accuracy.

There are some expressions where the model could correctly estimate that the rule-based method failed to parse. Since the expressions are user-generated, there are many free-form descriptions such as “$\overset{\text{2 servings (6 small portions)}}{\text{2 回分 (小サイズ 6 個)}}$” and “$\overset{\text{about 1 serving of salad for 4 people}}{\text{約 4 人分のサラダ 1 回分}}$”. These expressions show the servings by two different criteria, so the rule-based method of calculating the average of the two numbers fails. The fine-tuned generation model could parse these expressions even if the training dataset is constructed from the rule-based method and contains misparsed examples.

6. Conclusion

We proposed the JFC-Recipe dataset, a novel Japanese recipe dataset for nutrient estimation from recipes. We provided a data collection procedure, annotation process, statistics, and inter-annotator agreements. In experiments, we formulated food item estimation and quantity estimation and investigated estimation performance with pre-trained language model-based methods. Future work include scaling the dataset, transferring knowledge to image-based calorie estimation, addressing long-tailed labels, more accurate amount estimation for ambiguous unit expressions utilizing LLM’s web-scale knowledge.

7. Acknowledgments

In this paper, we used “Cookpad dataset” provided by Cookpad Inc. via IDR Dataset Service of National Institute of Informatics. This work was supported by JSPS KAKENHI Grant Number JP23K25247 and JST NEXUS, Japan Grant Number JPMJNX25C9. We would like to thank the anonymous reviewers for their insightful comments.

8. Ethics Statement

The JFC-Recipe dataset is based on the Cookpad dataset, which is collected from user-generated recipes. The Cookpad dataset does not contain any personally identifiable information, and our annotation also does not include such information.

The JFC-Recipe is based on recipes written in Japanese on Japanese web service. Thus, the dataset may contain cultural bias specific to Japan.

9. Data Availability Statement

Our annotation dataset is available at our project page[14] under CC BY-NC 4.0 license. The data contains recipe IDs, food item annotations, and quantity annotations. Recipes can be obtained from the Cookpad dataset by using the recipe IDs.

10. Bibliographical References

Ashkan Afshin, Patrick John Sur, Kairsten A Fay, Leslie Cornaby, Giannina Ferrara, Joseph S Salama, Erin C Mullany, Kalkidan Hassen Abate, Cristiana Abbafati, Zegeye Abebe, et al. 2019. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. The lancet, 393(10184):1958–1972.

Dhaivat J. Bhatt, Seyed Ahmad Abdollahpouri Hosseini, Federico Fancellu, and Afsaneh Fazly. 2024. End-to-end parsing of procedural text into flow graphs. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), pages 5833–5842. ELRA and ICCL.

Antoine Bosselut, Omer Levy, Ari Holtzman, Corin Ennis, Dieter Fox, and Yejin Choi. 2018. Simulating action dynamics with neural process networks. In Proceedings of

---

[14] `https://github.com/ku-lsta/JFC-Recipe`

the 2018 International Conference on Learning Representations (ICLR).

Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. 2021. Autoregressive entity retrieval. In Proceedings of the 2021 International Conference on Learning Representations (ICLR).

Yong Cao, Yova Kementchedjhieva, Ruixiang Cui, Antonia Karamolegkou, Li Zhou, Megan Dare, Lucia Donatelli, and Daniel Hershcovich. 2024. Cultural adaptation of recipes. Transactions of the Association for Computational Linguistics (TACL), 12:80–99.

Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. 2018. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), page 35–44. Association for Computing Machinery.

Jing-Jing Chen, Chong-Wah Ngo, Fu-Li Feng, and Tat-Seng Chua. 2018. Deep understanding of cooking procedure for cross-modal recipe retrieval. In Proceedings of the 26th ACM International Conference on Multimedia (ACM Multimedia), page 1020–1028. Association for Computing Machinery.

Donghee Choi, Mogan Gim, Samy Badreddine, Hajung Kim, Donghyeon Park, and Jaewoo Kang. 2023. Kitchenscale: Learning to predict ingredient quantities from recipe contexts. Expert Systems with Applications, 224:120041.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1):37–46.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (NAACL), pages 4171–4186. Association for Computational Linguistics.

Darius Feher, Faridz Ibrahim, Zhuyan Cheng, Viktor Schlegel, Tom Maidment, James Bagshaw, and Riza Batista-Navarro. 2023. Few-shot entity linking of food names. Information Processing & Management, 60(5):103463.

Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. Statistical methods for rates and proportions. john wiley & sons.

Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. 2024. Continual pre-training for cross-lingual LLM adaptation: Enhancing japanese language capabilities. In Proceedings of the First Conference on Language Modeling (COLM).

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models.

Jun Harashima, Michiaki Ariga, Kenta Murata, and Masayuki Ioki. 2016. A large-scale recipe and meal data collection as infrastructure for food research. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA).

Jun Harashima, Makoto Hiramatsu, and Satoshi Sanjo. 2020. Calorie estimation in a real-world recipe service. Proceedings of the AAAI Conference on Artificial Intelligence, 34(08):13306–13313.

Jun Harashima and Yoshiaki Yamada. 2018. Two-step validation in character-based ingredient normalization. In Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, CEA/MADiMa '18, page 29–32. Association for Computing Machinery.

Atsushi Hashimoto, Tetsuro Sasada, Yoko Yamakata, Shinsuke Mori, and Michihiko Minoh. 2014. Kusk dataset: toward a direct understanding of recipe text and human cooking activity. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct, page 583–588. Association for Computing Machinery.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In Proceedings of the 2022 International Conference on Learning Representations (ICLR).

Jermsak Jermsurawong and Nizar Habash. 2015. Predicting the structure of cooking recipes. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 781–786. Association for Computational Linguistics.

Chloé Kiddon, Ganesa Thandavam Ponnuraj, Luke Zettlemoyer, and Yejin Choi. 2015. Mise en place: Unsupervised interpretation of instructional recipes. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 982–992. Association for Computational Linguistics.

Najoung Kim and Sebastian Schuster. 2023. Entity tracking in language models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), pages 3835–3855. Association for Computational Linguistics.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In Proceedings of the 2020 Advances in Neural Information Processing Systems (NeurIPS), volume 33, pages 9459–9474. Curran Associates, Inc.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In Proceedings of the 2019 International Conference on Learning Representations (ICLR).

Koki Maeda, Tosho Hirasawa, Atsushi Hashimoto, Jun Harashima, Leszek Rybicki, Yusuke Fukasawa, and Yoshitaka Ushiku. 2024. Com kitchens: An unedited overhead-view video dataset as a vision-language benchmark. In Proceedings of the 2024 European Computer Vision Association (ECCV), pages 123–140.

Hirokuni Maeta, Tetsuro Sasada, and Shinsuke Mori. 2015. A framework for procedural text understanding. In Proceedings of the 14th International Conference on Parsing Technologies (IWPT), pages 50–60. Association for Computational Linguistics.

Javier Marın, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2021. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1):187–203.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pages 1906–1919. Association for Computational Linguistics.

Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P. Murphy. 2015. Im2calories: Towards an automated mobile vision food diary. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Shinsuke Mori, Hirokuni Maeta, Yoko Yamakata, and Tetsuro Sasada. 2014. Flow graph corpus from recipe texts. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC), pages 2370–2377. European Language Resources Association (ELRA).

Taichi Nishimura, Atsushi Hashimoto, and Shinsuke Mori. 2019. Procedural text generation from a photo sequence. In Proceedings of the 12th International Conference on Natural Language Generation (INLG), pages 409–414. Association for Computational Linguistics.

Taichi Nishimura, Suzushi Tomori, Hayato Hashimoto, Atsushi Hashimoto, Yoko Yamakata, Jun Harashima, Yoshitaka Ushiku, and Shinsuke Mori. 2020. Visual grounding annotation of recipe flow graph. In Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC), pages 4275–4284. European Language Resources Association.

Naoaki Okazaki, Kakeru Hattori, Hirai Shota, Hiroki Iida, Masanari Ohi, Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Rio Yokota, and Sakae Mizuki. 2024. Building a large japanese web corpus for large language models. In Proceedings of the First Conference on Language Modeling (COLM).

Liang-Ming Pan, Jingjing Chen, Jianlong Wu, Shaoteng Liu, Chong-Wah Ngo, Min-Yen Kan, Yugang Jiang, and Tat-Seng Chua. 2020. Multimodal cooking workflow construction for food recipes. In Proceedings of the 28th ACM International Conference on Multimedia (ACM Multimedia), page 1132–1141. Association for Computing Machinery.

Amaia Salvador, Michal Drozdzal, Xavier Giro-i Nieto, and Adriana Romero. 2019. Inverse cooking: Recipe generation from food images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Amaia Salvador, Erhan Gundogdu, Loris Bazzani, and Michael Donoser. 2021. Revamping cross-modal recipe retrieval with hierarchical transformers and self-supervised learning. In

Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15475–15484.

Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Takayuki Sato, Jun Harashima, and Mamoru Komachi. 2016. Japanese-English machine translation of recipe texts. In Proceedings of the 3rd Workshop on Asian Translation (WAT), pages 58–67. The COLING 2016 Organizing Committee.

Keisuke Shirai, Atsushi Hashimoto, Taichi Nishimura, Hirotaka Kameko, Shuhei Kurita, Yoshitaka Ushiku, and Shinsuke Mori. 2022. Visual recipe flow: A dataset for learning visual state changes of objects with recipe flows. In Proceedings of the 29th International Conference on Computational Linguistics (COLING), pages 3570–3577. International Committee on Computational Linguistics.

Niket Tandon, Keisuke Sakaguchi, Bhavana Dalvi, Dheeraj Rajagopal, Peter Clark, Michal Guerquin, Kyle Richardson, and Eduard Hovy. 2020. A dataset for tracking entities in open domain procedural text. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 6408–6417. Association for Computational Linguistics.

Quin Thames, Arjun Karpur, Wade Norris, Fangting Xia, Liviu Panait, Tobias Weyand, and Jack Sim. 2021. Nutrition5k: Towards automatic nutritional understanding of generic food. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8903–8911.

Atsushi Ushiku, Hayato Hashimoto, Atsushi Hashimoto, and Shinsuke Mori. 2017. Procedural text generation from an execution video. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP), pages 326–335. Asian Federation of Natural Language Processing.

Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, Griffin Thomas Adams, Jeremy Howard, and Iacopo Poli. 2025. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), pages 2526–2547. Association for Computational Linguistics.

Walter Willett, Johan Rockström, Brent Loken, Marco Springmann, Tim Lang, Sonja Vermeulen, Tara Garnett, David Tilman, Fabrice DeClerck, Amanda Wood, et al. 2019. Food in the anthropocene: the eat–lancet commission on healthy diets from sustainable food systems. The lancet, 393(10170):447–492.

Yoko Yamakata, Shinsuke Mori, and John Carroll. 2020. English recipe flow graph corpus. In Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC), pages 5187–5194. European Language Resources Association.

Luowei Zhou, Chenliang Xu, and Jason J Corso. 2018. Towards automatic learning of procedures from web instructional videos. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI), pages 7590–7598.