Defining and Detecting Incomplete Ingredient Descriptions in Cooking Recipes

Masatoshi Tsuchiya and Daigo Kohno
Toyohashi University of Technology
1–1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan
{tsuchiya, kohno}@is.cs.tut.ac.jp

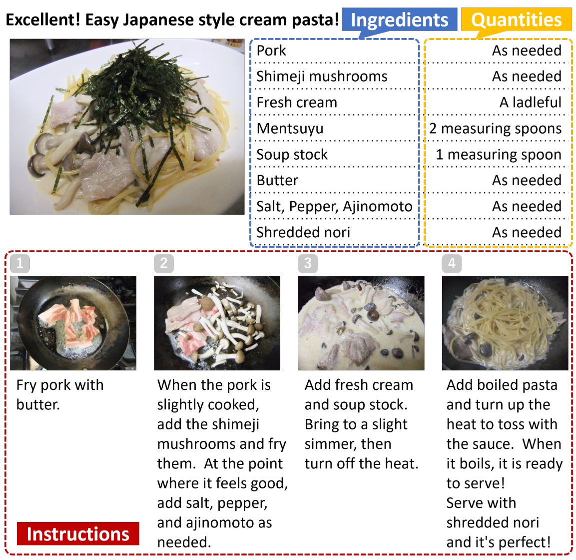

Figure 1: An example of an incomplete recipe, which is translated from https://cookpad.com/recipe/1190947 with our notes.

Abstract

This paper introduces the concept of complete ingredient descriptions as a feature of cooking recipes. It argues that a recipe provides a complete description of its ingredients if it satisfies two conditions: all ingredients listed in its ingredient list are mentioned in its cooking instructions, and all ingredients appearing in its cooking instructions are listed in its ingredient list. A new ingredient dictionary is constructed, and we show how it can be employed to determine whether a recipe has a complete ingredient description. Using this dictionary, it is experimentally demonstrated that at least 9.0% of a large dataset of user-generated recipes have incomplete ingredient descriptions, illustrating the need to consider such incomplete descriptions when processing recipes.

1 Introduction

A procedural document describes a sequence of instructions that must be followed to to achieve a specific goal, such as the process of assembling parts into a finished product or charting a route from one point to another (Delpech and Saint-Dizier, 2008). Reproducibility, which refers to whether the specific goal of a sequence of instructions can be achieved again, is a crucial requirement in procedural documents. For instance, to evaluate the effectiveness of a tool that assists human authors in writing procedural documents, Colineau et al. (2002) examined whether those who read the documents written with the tool were able to reproduce the goals described in those documents. To ensure reproducibility, complete documentation of instruction sequences is important, as is often pointed out in relation to the reproducibility of academic research (Glasziou et al., 2014; Beam et al., 2020; Storks et al., 2023).

A cooking recipe is a typical procedural document that describes a sequence of cooking instructions for making a dish from the listed ingredients, and it has attracted attention as a model domain for various NLP tasks (Momouchi, 1980; Tasse and Smith, 2008; Mori et al., 2014; Jermsurawong and Habash, 2015; Jiang et al., 2020). However, previous studies have uniformly treated all cooking recipes as complete procedural documents without considering the possibility that the target recipes do not provide complete documentation to reproduce their dishes. To address this problem, it is necessary to distinguish between complete and incomplete recipes.

The three basic elements of a recipe are the ingredients, their quantities and the instructions. Therefore, these three elements are considered to be the factors that could make a recipe unreadable. The first factor is an incomplete description of the ingredients. It is a situation in which an ingredient listed in the ingredient list is not mentioned in the instructions, or conversely, an ingredient appearing in the instructions is not listed in the ingredient list. For example, in Figure 1, “mentsuyu,” which is listed in the ingredient list, is not mentioned in the instructions, so it is unclear how to use “mentsuyu” in making this dish. In addition, although “pasta” appears in the fourth step of the instructions in Figure 1, it is not listed in the ingredient list. Such incomplete descriptions of the ingredients make the recipe unreadable.

The second factor is the incomplete description of quantities. Incomplete description of the quantities means that specific quantities are not described in the ingredient list or the instructions. For example, in Figure 1, most quantities of the ingredient list are specified as “as needed”. In addition, the amount of “cream” is specified as “a ladleful”, but unlike measuring spoons, the amount of “ladle” is not standardized. These ambiguous phrases make it difficult for readers to determine the accurate ingredient measurements.

The third factor is the incomplete description of instructions. An incomplete description of the instructions is a situation in which necessary actions are not described or an explanation of the aspect of the action is lacking. For example, the second instruction in Figure 1 uses the ambiguous expression “at the point where it feels good” as the end condition of the action “fry”, making it difficult to understand how long to fry the ingredients. There is the mention to “boiled pasta” in the fourth step of the instructions, but no action boiling pasta appears in the steps until the fourth step.

Of these three factors, the incomplete descriptions of quantities and instructions can vary greatly depending on the knowledge and skill level of the reader. The previously mentioned phrases, such as “as needed” and “at the point where it feels good,” are expected to be easily understood by a reader familiar with the recipe shown in Figure 1. Therefore, we limit the scope of this paper to the incomplete description of ingredients.

The main contributions of this paper are twofold:

1. This paper defines criteria for complete ingredient descriptions by comparing professionally reviewed recipes and user-generated recipes that were not reviewed by an expert or a third party.

2. Using a dictionary-based method, an experiment shows that at least 9.0% of a large dataset of user-generated recipes have incomplete ingredient descriptions, illustrating the need to consider such incomplete descriptions when processing recipes.

2 Related Works

There are two research streams related to this study: procedural document structure analysis and document completeness metrics.

The analysis and representation of recipe structure has been the subject of many previous studies. Tasse and Smith (2008) proposed a formal language based on first-order logic to annotate a small dataset of well-written recipes consisting of complete ingredient lists and instructions. Jermsurawong and Habash (2015) introduced a dependency tree format and empirically validated its applicability using the same dataset. These studies considered both the ingredient lists and the instructions but ignored the diversity of user-generated recipe expressions and the challenge of representing incomplete instruction sequences.

Momouchi (1980) proposed a flow graph for representing procedural documents, using a rule-based approach on a limited set of well-written recipes consisting of complete instructions. Mori et al. (2014) introduced an alternative flow graph format for representing 266 user-generated, well-written recipes. Jiang et al. (2020) adopted the frame-semantic representation of PropBank (Kingsbury and Palmer, 2002) for representing recipes. However, these studies ignored ingredient lists and did not address the challenge of representing incomplete instruction sequences.

In addition, Dalvi et al. (2019) and Zhang et al. (2023) focused their research on the dependency relationships between entities and events in generic procedural documents. Dalvi et al. (2019) employed a machine learning approach to predict dependencies between events and their purposes. Zhang et al. (2023) employed a LLM-based approach to determine which event changes the state of an entity that appears in a procedure document. Neither of them considered the case where an incomplete procedural document cannot has no appropriate dependency structure.

Document summarization is a task that removes redundant fragments from an input document and generates a summary of a given length. To accomplish this task, various metrics have been proposed to evaluate the quality of generated summaries. Nenkova et al. (2007) introduced a metric that assesses the quality of automatically gen-

Category	Subcategory	Definition	Example
Equipment	Reusable	Repeatedly usable cooking equipment	a pan, a bowl, a measuring cup
	Disposable	Equipment that can be used only once	a bamboo skewer, plastic wrap
Ingredient	Foodstuff	Foodstuffs actually ingested	a tomato, pork, flour (for dough)
	Auxiliary Foodstuff	Foodstuffs used in the cooking instructions but not ingested	kelp (for soaking), oil (for greasing), flour (for dusting)
Water		Water	water, lukewarm water

Table 1: Categories of items in the ingredient lists in the Cookpad dataset. Unlike professionally reviewed recipes, the ingredient lists of user-generated recipes contain a large variety of items. Since many recipes include disposables in their ingredient lists, but few recipes include reusable utensils, the "equipment" category is divided into two subcategories. Since main foodstuffs are definitely listed in the ingredient lists, but auxiliary foodstuffs are often missing, the "ingredient" category also needs to be divided. The "water" category is a special class that emerged as necessary during the analysis recounted in Section 3.2.

erated summaries by measuring the coverage of hierarchical content units derived from human-authored, gold-standard summaries. Takamura and Okumura (2009) further formalized document summarization as the problem of maximizing the coverage of conceptual units (Filatova and Hatzivassiloglou, 2004). According to these metrics, a generated summary can be considered a complete representation of the input document if it covers all the semantic units present in the original text. The contribution of this paper can be viewed as the development of a scale that can measure the completeness of cooking recipes by treating ingredients appearing in ingredient lists or instructions as semantic units.

3 The Criteria for a Complete Ingredient Description

In this paper, we define a recipe as having a complete ingredient description if it meets the following four conditions:

1. All ingredients listed in the ingredient list must be referenced in the instructions.
2. The ingredient list must include all ingredients referenced in the instructions, unless they are explicitly indicated as "optional" or "unlisted."
3. Equipment, whether reusable or disposable, need not be included in the ingredient list.
4. Items belonging to the "water" category may or may not be included in the ingredient list.

The subsequent sections detail the process that led to the establishment of these four conditions, through a comparison of professionally reviewed recipes and user-generated recipes that were not reviewed by an expert or third party.

3.1 Items on the Ingredient Lists of User-Generated Recipes

This section discusses which items our criteria for a complete ingredient description need to cover, with reference to the ingredient lists of user-generated recipes that have not been reviewed by an expert or a third party.

This paper investigates the Cookpad dataset (Harashima et al., 2016), which consists of a large number of user-generated recipes posted on a web platform designed for direct sharing among recipe creators. Due to the nature of this platform, the recipes in the Cookpad dataset were not reviewed by experts or third parties. The only guideline this platform provides regarding the ingredient list is that "ingredients and seasonings should be listed in the ingredient list," which is minimal and open to interpretation[1]. Consequently, there is substantial variation in the items included in the ingredient lists across different recipes.

Table 1 summarizes our investigation of the items listed in the ingredient lists in the Cookpad dataset. When supposing an item, the categorization of Table 1 focuses on whether the judgment of whether or not it should be listed in the ingredient list differs among recipe creators. First, for equipments used in the instructions, the reusable subcategory and the disposable subcategory are established, because a large difference was observed between reusable equipments, such as a pan and a measuring cup, and equipments that can be used only once, such as a bamboo skewer or cooking sheet. Many recipes include disposables in their ingredient lists, whereas only a few recipes include reusables. Next, a large difference was observed between the foodstuffs actually ingested and those

[1] The original guideline, written in Japanese, can be found at https://cookpad.com/recipe/post/help.

# of TV programs	11
# of recipes	2320
# of items on the ingredient lists	21851
# or instruction steps	10786
# of recipe creators	296

Table 2: Statistics from the NHK dataset, which consists of professionally reviewed recipes collected from the site on January 30, 2021.

not actually ingested. For example, most recipes include flour for dough in their ingredient lists, whereas many recipes do not include flour for dusting. For this reason, the ingredient category is divided into the foodstuff subcategory and the auxiliary foodstuff subcategory. The foodstuff subcategory is defined as foodstuffs that are actually ingested, and the auxiliary foodstuff subcategory is defined as foodstuffs that are used in the instructions but are not actually ingested. Finally, water is a special category that became necessary in the analysis described below.

3.2 Items on the Ingredient Lists of Professionally Reviewed Recipes

This section isolates the criteria that a complete ingredient description must meet by investigating professionally reviewed recipes.

Table 2 presents a statistical breakdown of professionally reviewed recipes collected from the website[2] of a set of culinary TV programs produced by the Japan Broadcasting Corporation. These recipes (hereafter referred to as the *NHK dataset*) were authored by culinary experts and reviewed by program production professionals. Therefore, unlike the user-generated recipes in the Cookpad dataset, the recipes in the NHK dataset are considered to be highly quality-controlled. This paper assumes that the review standards implicitly applied to the recipes in the NHK dataset correspond to the criteria that a complete ingredient description must satisfy.

Unfortunately, since the review standards of the NHK dataset are not explicitly disclosed, it is necessary to derive them from the actual recipes in the dataset, as discussed in the following paragraphs. Our investigation to derive these standards from the NHK dataset consists of two steps. The first step is to derive the review standards from the relationship between the ingredient lists and the instructions. This step involves a manual investigation of the mappings between the items listed in the ingredient lists and those mentioned in the instructions. The second step is to derive the review standards from the items listed in the ingredient lists. This step involves a manual investigation of these items, categorized according to Table 1.

[2]https://www.nhk.or.jp/lifestyle/recipe/

Subcategory	Example	# in instr.	# in ingrd.
Reusable	a pan	930	0
	a pot	811	0
Disposable	plastic wrap	287	0
	a bamboo skewer	98	0
Foodstuff	a tomato	305	305
	pork	238	238
	kelp	23	23
	starch	191	191
Auxiliary Foodstuff	kelp	27	27
	starch	8	8
Water	water	863	415

Table 3: Occurrences of typical items in the instructions and ingredient lists in the NHK dataset. Our manual investigation of "kelp" and "starch," which can be used as both main and auxiliary foodstuffs, revealed that they must be listed regardless of their subcategories. In other words, all ingredients must be listed in the ingredient list, but not all equipment needs to be listed.

As the first step, we manually investigated the mappings between the items listed in the ingredient lists and those appearing in the instructions for 500 recipes randomly sampled from the NHK dataset. From this investigation, two review standards were identified. First, all items listed in the ingredient lists must be mentioned in the instructions. In other words, the NHK dataset does not permit any omissions where an item listed in the ingredient list is absent from the instructions, as illustrated by the "mentsuyu" example in Figure 1. Second, the ingredient list must include all ingredients referenced in the instructions. If an item not listed in the ingredient list is mentioned in the instructions, it is explicitly noted with phrases such as "optional" or "unlisted," as in the following example:

> When the dough has doubled in size from 1.5 to 2 times, open the lid and dust with flour (unlisted)[3].

As the second step, we manually examined whether the typical items for the subcategories shown in Table 1 are listed in the ingredient lists

[3]The example is translated from https://www.nhk.or.jp/lifestyle/recipe/detail/500360.html.

Examples in ingredient lists	Examples in instructions	Description	Frequency
a tomato	Cut a tomato	The exact same string is used.	2320 (65.5%)
stew blend	Add stew mix	A different name for the same ingredient is used.	329 (9.3%)
a can of tomatoes	Boil tomatoes	The pre-processed name refers to the processed one.	98 (2.8%)
an onion, a tomato	Cut vegetables	A class name that covers several ingredients is used.	269 (7.6%)
♠pork, ♠an onion	Fry ♠	A special symbol is used in the ingredient list.	491 (13.9%)
an onion, a salmon	Cut all	An explanation specifies a subset of the ingredients.	324 (9.1%)

Table 4: Variations in ingredient mappings in the 500 recipes randomly sampled from the Cookpad dataset. The first row shows that simple string matching can identify only 65.5% of the mappings.

of the NHK dataset. Two additional review standards were identified from the results of this investigation, as summarized in Table 3. The first review standard is that equipment, whether reusable or disposable, does not need to be listed in the ingredient lists. For example, while "a pan," categorized as reusable, appears in the instructions of 930 recipes in the NHK dataset, it is never listed in the corresponding ingredient lists. Similarly, "a pot" (reusable), "plastic wrap," and "a bamboo skewer" (both disposable) are mentioned in the instructions but are not listed in the ingredient lists. This indicates that the NHK dataset follows a review standard where equipment, whether reusable or disposable, is consistently omitted from the ingredient lists.

The second review standard identified from Table 3 is that every ingredient must be listed in the ingredient list, regardless of whether it is used as a foodstuff or an auxiliary foodstuff. For example, "a tomato" and "pork," both consistently categorized under the foodstuff subcategory, are always included in the ingredient lists, as shown in Table 3. In contrast, ingredients that can serve either as a foodstuff or as an auxiliary foodstuff require manual inspection to determine their subcategories. For instance, "kelp" may be used as a foodstuff in some recipes and as an auxiliary foodstuff for soaking and discarding in others. Thus, simply counting occurrences of "kelp" does not clarify its subcategory. Our manual inspection of recipes where "kelp" appears reveals that it is used as a foodstuff in 25 recipes and as an auxiliary foodstuff in 22 recipes; in both cases, "kelp" is listed in the ingredient lists, regardless of its subcategory. Similarly, "starch" and "flour" are used as either foodstuffs or auxiliary foodstuffs in various recipes. Our manual inspection of these ingredients also reveals that they are consistently listed in the ingredient lists when they are mentioned in the instructions (except when there are explicit notes). Based on these observations, it can be assumed that the NHK dataset adopts the review standard that every ingredient must be listed in the ingredient lists, regardless of whether it is used as a foodstuff or an auxiliary foodstuff.

The situation with items that belong to the water category is very different from those that belong to the equipment category or the ingredient category. Even if "water" appears in the instructions, "water" may or may not be listed in the ingredient list. The manual inspection of 100 recipes randomly sampled from recipes where "water" appears in their instructions revealed that 42 recipes used "water" as a foodstuff and 62 recipes used "water" as an auxiliary foodstuff[4]. Therefore, it can be assumed that there is no review standard for items belonging to the water category in the NHK dataset.

Through these observations, the four criteria stated at the beginning of Section 3 are obtained.

4 Analysis of Incomplete Ingredient Descriptions

4.1 Detection of Incomplete Ingredient Descriptions

To illustrate the need to watch out for incomplete ingredient descriptions in recipes, this section discusses the proportion of incomplete ingredient descriptions in the Cookpad dataset. Based on the criteria described in Section 3, the automatic detection of incomplete ingredient descriptions in recipes requires the mapping between ingredients listed in ingredient lists and those appearing in instructions. Since there are diverse ways to express ingredients, especially in user-generated recipes, a mapping method is needed that can handle such diverse expressions.

Table 4 shows our manually annotated results of the mappings between ingredients listed in the ingredient lists and those appearing in the instructions of the 500 recipes randomly sampled from the Cookpad dataset. Since only 65.5% of the

[4] 4 recipes used "water" as both subcategories.

All ingredients mentioned	952k (89.5%)
All ingredients listed	953k (89.6%)
Complete ingredient descriptions	860k (80.8%)
Total	1064k

Table 5: Detection of complete ingredient descriptions. 89.5% of the recipes are complete in terms of instructions since their instructions mention all the ingredients listed in their ingredient lists. 89.6% of the recipes are complete in terms of ingredient lists. 80.8% of the recipes satisfy both criteria, resulting in complete ingredient descriptions.

mappings could be identified using simple string matching, a dictionary-based mapping method is necessary and seemed feasible as a way to identify the remaining mappings that involve string modifications. Note that the mappings in the last row of Table 4 were excluded from our scope because there were too many different expressions in the last row to identify.

Based on the above observation, the new dictionary was constructed by merging the two existing dictionaries (Nanba et al., 2014; Kiyomaru et al., 2018), and by manually collecting ingredients that appear 10 or more times in the Cookpad dataset. Although our dictionary achieved 99.0% coverage, which is higher than the 97.0% and 94.8% coverage rates of the existing ones, further improvement in coverage remains a difficult challenge due to the diversity of expressions.

Table 5 presents the experimental results, revealing that 80.8% of the recipes have complete ingredient descriptions. The remaining 19.2% of recipes, while potentially incomplete, are not guaranteed to be so due to our dictionary's limited coverage of ingredients in the Cookpad dataset. Our manual examination of the 200 recipes randomly sampled from this 19.2% confirmed that 94 recipes were indeed incomplete. Consequently, the dictionary-based method achieved an accuracy rate of 89.8% in detecting complete or incomplete ingredient descriptions, indicating that 9.0% of all recipes could be considered incomplete.

4.2 Relation to User Feedback

This section discusses the relationship between incomplete ingredient descriptions and user feedback posted by the users of the recipe-sharing site to express their impressions of the recipes. While 47.0% of recipes with complete ingredient descriptions received feedback, 41.7% of recipes with incomplete ingredient descriptions received feedback. Therefore, recipes with complete and incomplete ingredient descriptions have statistically significantly different probabilities of receiving feedback, but the effect size is not large.

We manually examined recipes with incomplete ingredient descriptions and feedback. Since most users of the recipe-sharing site are not novices, they can effectively reproduce dishes even with incomplete descriptions if the dishes are appealing. As a result, the number of feedback for a recipe is more indicative of its appeal than its reproducibility, especially for experienced cooks. This raises the possibility that the concept of complete ingredient descriptions may be only one element influencing reproducibility, which requires further research.

5 Conclusion

This paper defined the criteria of complete ingredient descriptions in terms of the mapping between ingredients listed in the ingredient lists and those appearing in the instructions, through comparing professionally reviewed recipes and user-generated recipes that are not reviewed by an expert or a third party. The new ingredient dictionary was constructed by merging the two existing dictionaries and collecting ingredients from the Cookpad dataset, which is a large dataset of user-generated recipes, and was employed to determine recipes with complete ingredient descriptions. The experiment on the Cookpad dataset showed that there were at least 9.0% of recipes with incomplete ingredient descriptions, illustrating the need to consider incomplete ingredient descriptions while processing user-generated recipes.

In the future, we plan to go beyond the problem of complete ingredient description to consider the definition of a complete recipe. In addition, the relationship between completeness and reproducibility will be analyzed in more detail.

References

Andrew L. Beam, Arjun K. Manrai, and Marzyeh Ghassemi. 2020. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA*, 323(4):305–306.

Nathalie Colineau, Cecile Paris, and Keith Vander Linden. 2002. An evaluation of procedural instructional text. In *Proceedings of the International Natural*

Language Generation Conference, pages 128–135. Association for Computational Linguistics.

Bhavana Dalvi, Niket Tandon, Antoine Bosselut, Wen-tau Yih, and Peter Clark. 2019. Everything happens for a reason: Discovering the purpose of actions in procedural text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4496–4505. Association for Computational Linguistics.

Estelle Delpech and Patrick Saint-Dizier. 2008. Investigating the structure of procedural texts for answering how-to questions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 46–51. European Language Resources Association (ELRA).

Elena Filatova and Vasileios Hatzivassiloglou. 2004. A formal model for information selection in multi-sentence text extraction. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 397–403. COLING.

Paul Glasziou, Douglas G. Altman, Patrick Bossuyt, Isabelle Boutron, Mike Clarke, Steven Julious, Susan Michie, David Moher, and Elizabeth Wager. 2014. Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, 383(9913):267–276.

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 2016)*, pages 2455–2459. European Language Resources Association (ELRA).

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*, pages 781–786. Association for Computational Linguistics.

Yiwei Jiang, Klim Zaporojets, Johannes Deleu, Thomas Demeester, and Chris Develder. 2020. Recipe instruction semantics corpus (RISeC): Resolving semantic structure and zero anaphora in recipes. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 821–826. Association for Computational Linguistics.

Paul Kingsbury and Martha Palmer. 2002. From Tree-Bank to PropBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. European Language Resources Association (ELRA).

Hirokazu Kiyomaru, Sadao Kurohashi, Mitsuru Endoh, and Katsuyoshi Yamagami. 2018. Construction of basic cooking knowledge base based on cooking recipes and crowdsourcing. In *Proceedings of the 24th Annuaal Meeting of The Association for Natural Language Processing*, pages 662–665. (in Japanese).

Yoshio Momouchi. 1980. Control structures for actions in procedural texts and PT-chart. In *COLING 1980 Volume 1: The 8th International Conference on Computational Linguistics*.

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'14)*, pages 2370–2377. European Language Resources Association (ELRA).

Hidetsugu Nanba, Yoko Doi, Miho Tsujita, Toshiyuki Takezawa, and Kazutoshi Sumiya. 2014. Construction of a cooking ontology from cooking recipes and patents. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 507–516. Association for Computing Machinery.

Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. 2007. The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Trans. Speech Lang. Process.*, 4(2):4–27.

Shane Storks, Keunwoo Yu, Ziqiao Ma, and Joyce Chai. 2023. NLP reproducibility for all: Understanding experiences of beginners. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10199–10219. Association for Computational Linguistics.

Hiroya Takamura and Manabu Okumura. 2009. Text summarization model based on maximum coverage problem and its variant. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 781–789. Association for Computational Linguistics.

Dan Tasse and Noah A Smith. 2008. SOUR CREAM: Toward semantic processing of recipes. Technical Report CMU–LTI08–005, Carnegie Mellon University.

Li Zhang, Hainiu Xu, Yue Yang, Shuyan Zhou, Weiqiu You, Manni Arora, and Chris Callison-Burch. 2023. Causal reasoning of entities and events in procedural texts. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 415–431. Association for Computational Linguistics.