World Languages and Cultures Faculty Publications

World Languages & Cultures

2020

Regularity and Variation in Japanese Recipes: A Comparative Analysis of Cookbook, Online, and User-generated Sub-registers

Michiko Kaneyasu
Old Dominion University, mkaneyas@odu.edu

Minako Kuhara
University of Colorado, Boulder, minako.kuhara@colorado.edu

Regularity and variation in Japanese recipes:

A comparative analysis of cookbook, online, and user-generated sub-registers

Michiko Kaneyasu and Minako Kuhara

Old Dominion University | University of Colorado Boulder

Abstract

This paper investigates the similarities and differences between three sub-registers of Japanese recipe texts: cookbook recipes, online commercial recipes written/edited by professionals, and online user-generated recipes. Past studies on Japanese recipes do not distinguish different sub-registers, and they tend to focus on a single feature. The present study of the sub-registers examines a group of frequently appearing linguistic features and uncovers functional links between observed features and situational characteristics. The comparative perspective contributes to a more comprehensive understanding of the Japanese recipe language as well as universal and language-specific aspects of register variation. Shared traits among the three sub-registers are tied to the common topic of cooking and the central purpose of providing easy-to-follow food preparation instructions. Varied linguistic and textual features are motivated by different production circumstances, mediums, and relations among the participants. Professionally edited cookbook and online commercial recipes show a much higher uniformity in their grammatical features than unedited/self-edited user-generated recipes. Online sub-registers share a role of serving as a repository and reference center for numerous recipes and related information. Relationships among writers, readers, and other participants such as publishers and site organizers differ among all three sub-registers, resulting in some unique linguistic patterns.

1. Introduction

Register boundaries are not always clear, but recipes are often recognized as a standard and central case of a register (e.g., Zwicky & Zwicky 1982, Fischer 2013). We adopt the general definition of a register as the use of language in a particular social situation for a specific purpose (Ferguson 1994). Recipes are a

type of technical manual written for the general public to provide instructions for preparing food. Ones written for a general audience tend to be more explicit and detailed than those written for a group of professionals (Diemer 2013) or a smaller community of people (Cotter 1997). Before the internet era, most publicly available recipes were written by professionals (created by cooking experts, and written, revised, and edited by publishers) in the form of cookbooks. Since the late 1990s, more and more commercial recipes have been published on the internet; moreover, online recipes are increasingly produced by amateur writers. The trend justifies and calls for inclusion of online and user-generated varieties in a register study of recipes. Online commercial and user-generated recipes share a major purpose of providing cooking instructions with the cookbook recipes, but they differ in other aspects such as specific mediums and type of writers. In this study, we provide a comparative analysis of cookbook recipes, online commercial recipes, and online user-generated recipes as three sub-registers of a general recipe register.

As we will see in the subsequent sections, patterns of regularities and variations among the three sub-registers illustrate that the Japanese recipe register is more complex and flexible than it may first seem. Although the sub-registers exhibit lexical and organizational uniformity to a certain degree, they show dissimilarities both within and peripheral to the recipe proper. The study identifies three situational factors that explain those differences: production circumstances, medium, and relations among participants. In particular, editorial control, offline vs. online medium, and interactiveness between writers and readers are closely associated with the varied linguistic features found in the three sub-registers. Editorial control and interactiveness are related in that direct interaction between writers and readers only takes place in an environment with low or no editorial control. Writers under these circumstances tend to draw on colloquial language use, though there are also substantial individual differences.

Japanese recipes have been analyzed in a variety of perspectives, but the previous literature does not make a distinction between offline and online recipes or between commercial and user-generated ones. Moreover, past studies tend to focus on a specific aspect such as a linguistic form (Aoyama 1987 on the use of the topic particle *wa*, Ono 1990 on the medial verb forms, Moriya 1993 on transitive vs. intransitive verbs), translation issues between Japanese and English (Naganuma 2006, Yamakata, Carroll, & Mori 2017), and sociocultural comparison with other languages (Martinec 2003 on Japanese and English, Strauss, Chang, & Matsumoto 2018 on Japanese, Korean, and English). Among the past studies, Akiyama (2002) is the most exhaustive, covering lexical, structural, and some grammatical features, though the analysis is based on a few sample recipes. However, Akiyama's study takes a style perspective and does not consider functional motivations for the observed patterns. One major purpose of the present study is to give a more comprehensive description of the linguistic features of Japanese recipes, including

functional relationships between observed features and situational characteristics. A register is usually not identified by a single lexical or grammatical feature but by a group of frequently occurring linguistic features. Thus, it is important to look at a variety of linguistic features and examine their distribution relative to other (sub-)registers. The current comparative study of three recipe sub-registers aims to further our understanding of the relationships between various situational factors and linguistic features that may also be relevant to other types of written registers in Japanese.

The structure of the rest of the paper is as follows. In the next section, the theoretical and analytical framework of the present study is described. Section 3 presents the dataset and methodology. In Section 4, the situational and linguistic characteristics of the cookbook recipes are reported and given a functional interpretation. These characteristics are compared with features of online commercial and online user-generated texts in Section 5. Section 6 discusses some similarities and differences among the three sub-registers, and Section 7 summarizes the major findings and concludes the paper.

2. Theoretical and analytical framework

The present study follows the basic tenets and methods of Biber's (1994) and Biber and Conrad's (2009) register study approach (see also Basso 1974, Biber 1988, Halliday 1978, Hymes 1974). This approach takes a usage-based view of language that linguistic features are inseparable from actual language use and functions within particular situations of use (e.g., Du Bois 1987, Hopper 1998, Barlow & Kemmer 2000, Bybee 2006). In this view, if a particular lexical or grammatical feature appears much more frequently in Register A than Register B, it is because this feature functionally better fits and meets the situational needs or communicative purposes of Register A. A register study is comparative in nature; it identifies common or pervasive linguistic features in a single register, in comparison with other registers, and seeks a functional explanation that links the linguistic features and situational characteristics of the target register.

Register variation is ubiquitous in any language, but it has been given much less attention than regional and social dialect variation. Part of the reason seems to be the ambiguity of the central and related concepts of register, genre, and style. Instead of defining them as categories or varieties of texts, Biber and Conrad (2009) use these concepts as analytical instruments. The register perspective identifies frequently-occurring linguistic features and seeks a functional explanation based on the situational factors. The genre perspective describes conventional structures and linguistic features that may only occur once in a text (e.g., opening and closing of a letter) and link them to the situational contexts. The style perspective focuses on pervasive linguistic features associated with particular authors or time periods and interprets the associations with respect to aesthetic preferences. In the present study of recipe

texts, we employ the register perspective for the most part, but we also use the genre perspective when relevant.

Although register studies on languages other than English are scarce, the existing literature demonstrates both cross-linguistic similarities in the linguistic features of some registers and interesting cultural differences within comparable registers (e.g., Kittredge 1982, Kim & Biber 1994, Biber 1995, Biber, Davies, Jones, & Tracy-Ventura 2006). Only more empirical investigations on various languages can advance our understanding of the universal and culture-specific aspects of register variation and language use (Biber & Conrad 2009: 260). While the current study does not provide a cross-linguistic analysis, it aims to contribute to this undertaking by providing Japanese linguistic and cultural perspectives.

In this study, we employed Biber and Conrad's methodological framework developed for English (2009: 29–82). It involves the following basic steps: (1) description of the situational characteristics, (2) description of the linguistic characteristics, and (3) analysis of the functional or conventional associations between the situational and linguistic features. Major categories under situational characteristics include the type of addressors and addressees, relations between them, mode/medium, production circumstances, setting, communicative purposes, and topic. These categories are based on their survey of previous literature on register variation (e.g., Biber 1988, 1994, Crystal & Davy 1969, Halliday 1978, Hymes 1974, Basso 1974). Many types of linguistic features may be relevant in a register analysis, including vocabulary, word classes, verb features, noun phrases, adverbials, word order, and clause type. There is no exhaustive list of linguistic features that a register study can rely on; the type and number of features to analyze depend on the focus and goal of the study as well as practical considerations. In the next section, we describe the specific steps and methods we used in identifying the situational and linguistic characteristics of Japanese recipes in the present study.

3. Data and methods

The core dataset for the current study consists of 30 recipe texts each from cookbooks, online commercial recipe sites, and online user-generated recipe sites, totaling 90 texts. All commercial recipes used in this study are created and written by professionals, while all user-generated recipes are written and posted by amateur writers.[1] We recognize this is a small dataset, and the results we report in the subsequent sections are to be taken as preliminary. At the same time, because each dataset represents a highly specific sub-register, we believe the small number of texts is less problematic than if this were a study on a general recipe register. The recipes are collected from different food/dish categories and retyped by the researchers for further analysis. Sources include five cookbooks, three online commercial sites, and two online user-generated sites.[2]

In terms of the representativeness of linguistic features found in our data, both our informal examination of recipe texts in various cookbooks (and magazines) as well as reports from the previous studies on these types of recipes (see Section 1) indicate that the linguistic features found in cookbook recipe texts are highly uniform and consistent. Thus, we believe 30 cookbook recipe texts from five different cookbooks are sufficient to represent the typical linguistic features of this highly specialized and structured how-to text. Any deviations and variations from the features of cookbook recipes found in the other two sub-registers are only representative to the extent that the particular sources we have examined are well known and popular among the Japanese population (see Endnote 1). Some situational characteristics and associated linguistic features may not be applicable in other sources not included in our data.

In addition to the 90 recipe texts, we examined 40 news reporting texts as a sample of a different written register. News reporting texts share some basic characteristics with recipe texts: they are both learned expositions written for general public, and they both have relatively short text with clear beginning and end. The inclusion of the news reporting register for comparison makes the situational and linguistic characteristics of the recipe texts more visible and salient.[3] The news texts are collected from four major newspaper sites, covering various topics.[4] Table 1 shows the size of each corpus.

Table 1. Data size

	No. of text	Total No. of words	Average No. of words per text (range)
Cookbook recipes (Cookbook)	30	5,175	173 (89–540)
Online commercial recipes (Commercial)	30	11,503	383 (179–640)
Online user-generated recipes (User-generated)	30	18,714	624 (331–1,002)
News reports (News)	40	9,802	245 (81–903)

As mentioned earlier, we adopt Biber's (1994) and Biber and Conrad's (2009) register study approach in our investigation of the three sub-registers of recipe texts. We first identified characteristics of cookbook recipes, which were assumed to represent the most traditional and typical form of recipe language and text structure, by comparing their features with those of news reporting texts. In the absence of a publicly available large written language corpus in Japanese, news reports were chosen as a register for comparison for their availability and shared basic characteristics with recipe texts. The results of this analysis are reported in Section 4. We then compared the identified characteristics of the cookbook recipes with features found in online commercial recipes and online user-generated recipes, whose results are reported in Section 5.

The first step in the register study is describing the situational characteristics of the target register. The two researchers discussed the situational characteristics of a target (sub-)register based on their

experience, sample texts, and previous studies, and summarized the identified characteristics (see Sections 4.1 and 5.1). In the next step of describing typical or pervasive linguistic features, we consulted the list of potentially relevant linguistic features from previous register studies on English (Biber & Conrad 2009: 78–82) as well as previous linguistic or stylistic studies on Japanese recipes (i.e., Aoyama 1987, Ono 1990, Moriya 1993, Akiyama 2002), as we examined the texts, to decide which features to investigate further. As described in Sections 4 and 5, we mainly focused on lexicon, verb morphology, referential forms, and text structure. After identifying the linguistic features that are more prevalent in one (sub-)register or another, we sought functional explanations for these linguistic characteristics by exploring and interpreting the relationship between situational characteristics and frequent linguistic features. At this stage, it was necessary for us to go back to the list of situational characteristics and consider additional situational factors that may explain particular linguistic characteristics.

We used the following programs in our text analysis. Raw typed texts were tokenized using MeCab morphological analyzer (Kudo 2013) with IPADIC (IPA dictionary) (Asahara & Matsumoto 2003). MeCab with IPADIC was also used for POS tagging.[5] We used AntConc (Anthony 2014) for concordance and generating keyword lists based on log-likelihood (LL). To measure the lexical diversity of each text (Sections 4.2.1 and 5.2.1), we used Maas's a^2 (Maas 1972), $a^2 = (\log N - \log V)/(\log^2 N)$. Sentence length (Section 4.3.3) was measured in the number of words, using the tokenized texts and Excel LEN and SUBSTITUTE functions. The basic formula is: "=LEN(cell_name)-LEN(SUBSTITUTE(cell_name," ",""))+1". This formula counts the number of spaces within each cell and adds one to the total number as the number of words in a sentence is the number of spaces plus one. We included the punctuation marks ("、" and "。") in the calculation. Annotations for transitivity of verbs (Section 4.3.1 and 5.3.1), plain vs. polite verb forms (4.3.2 and 5.3.2), and major ingredients (4.4 and 5.4) were done manually, using POS-tagged concordance lines (for the first two annotations) and raw texts (for the third annotation). Ingredients were considered major if they were mentioned in the list of ingredients and they were not condiments, oil, seasonings, or ingredients for coating, sauce, or simple garnish.

4. Characteristics of cookbook recipes

In this section, we present the situational and linguistic characteristics of cookbook recipes (henceforth, Cookbook) in comparison with those of news reports (henceforth, News), and explain how they are functionally and/or conventionally associated with one another. To remain focused on the recipe register, characteristics of News will be mentioned only when necessary to highlight the characteristics of Cookbook.

4.1 Situational characteristics

Cookbook is written by professionals for unenumerated generic readers, with no interaction between the two parties. Texts are published in the printed medium, and they are carefully planned, revised, and edited. Cookbook deals with a topic of food in general and cooking in particular. The major communicative purpose of Cookbook is to provide procedural food preparation instructions. Although this was not part of our initial observation, it became clear through the analysis of linguistic features that the assumed setting of communication, that is, the fact that at least some readers would be glancing at the cooking instructions while cooking, and therefore they would have rather dispersed attention, is also an important situational characteristic that influences the linguistic and organizational properties of Cookbook.[6] In the following subsections (4.2–4.5), the linguistic features of Cookbook are described and explained vis-à-vis situational characteristics.

4.2 Vocabulary

4.2.1 Lexical diversity and categories

Lexical diversity of Cookbook is lower than that of News, as shown by a lower Maas' index score in Cookbook $M = 10.32$ ($SD = 5.38$) than in News $M = 14.42$ ($SD = 8.76$). This is statistically significant based on the independent t-test $t(68) = 2.258$, $p < .05$, $d = 0.56$. The higher lexical diversity in News is associated with a wider range of topics dealt in news reports such as local and national news, politics, business, international affairs, and sports. Cookbook only deals with the single topic of food or food preparation; thus it has a less varied lexicon.

Figure 1 shows that Cookbook has a higher proportion of verbs (1.4 times more), adverbs (3.5 times more), and adjectives (2.1 times more) than News. A closer look at adjective type in Cookbook reveals that nearly 60% are used adverbially; most frequent ones are *usuku* 'thinly' and *karuku* 'lightly.' The higher distribution of verbs, adverbs, and adjectives in Cookbook indicates its primary focus is on actions, that is, step-by-step procedures.
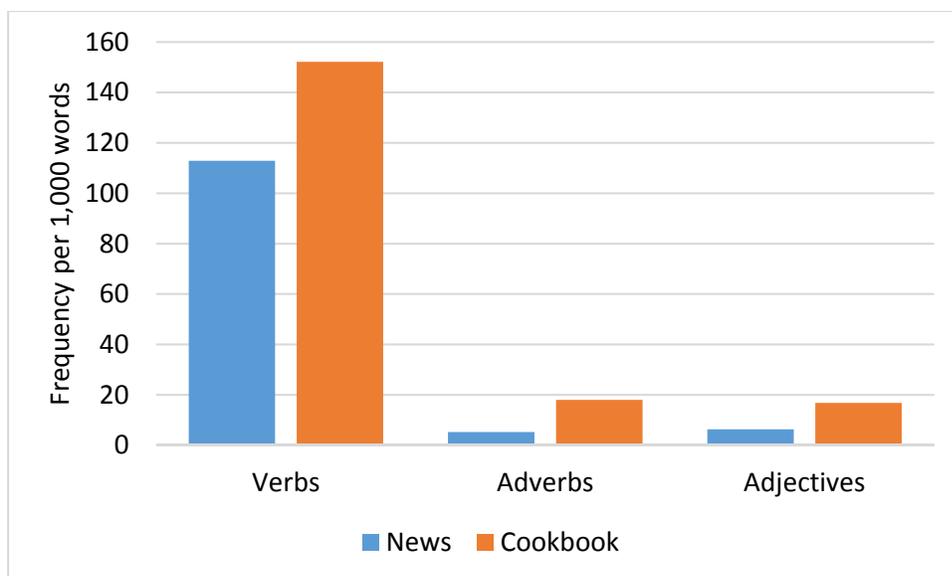
Figure 1. Frequency of some lexical categories in News and Cookbook (per 1,000 words)

4.2.2 Frequent words, keywords, and specialized words

The most frequent words and top keywords in Cookbook, as shown in Table 2, belong to one of the following four types: (1) particles with highly specific and limited functions [P], (2) cooking terminology [C], (3) quantity specifiers [Q], and (4) text organizers [T]. Three words that do not belong to any of the four categories, *shi* 'do,' *te* (medial verb form), and *masu* (polite verb suffix), will be discussed in Section 4.3.

Table 2. Most frequent words and keywords in Cookbook

	Most frequent words		Top keywords	
	Word	Meaning/function	Word	Meaning/function
1	*o*	Particle [P]	*oosaji*	Noun 'tablespoonful' [Q]
2	*ni*	Particle [P]	1	Quantity [Q]
3	*te*	*Te*-form	●	Bullet point [T]
4	*no*	Particle [P]	2	Procedural step* [T]
5	1	Quantity [Q]	*shio*	Noun 'salt' [C]
6	*wa*	Particle [P]	1	Procedural step [T]
7	2	Quantity [Q]	/	Fraction symbol [Q]
8	*oosaji*	Noun 'tablespoonful' [Q]	3	Procedural step [T]
9	*de*	Particle [P]	4	Quantity/serving size [Q]
10	4	Quantity/serving size [Q]	*shooyu*	Noun 'soy sauce' [C]
11	*to*	Particle [P]	*zairyoo*	Heading 'ingredients' [T]
12	*shi*	Verb 'do'	*ire*	Verb 'put in' [C]
13	●	Bullet point [T]	2	Quantity [Q]

14	*f(b)un*	Noun 'minute; for' [Q]	g	Unit 'gram' [Q]
15	2	Procedural step [T]	*tsukurikata*	Heading 'how to make' [T]
16	*shio*	Noun 'salt' [C]	4	Procedural step [T]
17	1	Procedural step [T]	*o*	Particle [P]
18	*ga*	Particle [P]	~	Tilde 'between' [Q]
19	/	Fraction symbol [Q]	*masu*	Polite verb suffix
20	*ire*	Verb 'put in' [C]	*kuwae*	Verb 'add' [C]

* The procedural step numbers include referential uses (see Section 4.4).

Particles are frequent in News as well, but their uses and functions are more limited in Cookbook. For example, the particle *ni* can generally be used for various functions, such as marking a destination or purpose of movement (e.g., *kodomo-tachi wa gakkou ni it-ta*. 'Children went <u>to</u> school.'), a specific time (e.g., *go-ji ni ie ni kaeru*. '(They) will come home <u>at</u> 5 o'clock.'), an indirect object of some action (e.g., *tomodachi ni hon o okut-ta* '(I) sent (my) friend a book.'), or an agent or source in passive or causative constructions (e.g., *inu ni kam-are-ta*. '(I) was bitten <u>by</u> a dog.'). In Cookbook, however, *ni* only has two main functions: (1) marking a noun as a location to which an ingredient is put, added, mixed, etc. (e.g., *furaipan ni abura o nesshi* 'heat oil <u>in</u> a frying pan') and (2) indicating a manner in which some action is performed, including shape, size, and heating level (e.g., *hanbun ni kiri* 'cut <u>in</u> half'). The particle *wa* also has a specific use, which will be described in Section 4.4.

There are also some specialized words in Cookbook: nominal compounds schematized as [shape]-*giri* '[shape]-cut' (e.g., *sen-giri* 'julienne strips,' *usu-giri* 'thin slice'), compound verbs (e.g., *maze-awaseru* 'mix together,' *mawashi-ireru* 'pour in a circular motion'), manner adverbs (e.g., *karitto* 'crisply,' *tororito* 'thickeningly'), and others (e.g., *kitsune-iro* 'golden brown, lit. fox color'). These are not register markers per se, but the topic-specific status of these terms is confirmed in the Balanced Corpus of Contemporary Written Japanese (BCCWJ).[7]

Prominent terms in Cookbook demonstrate that the texts center on actions (i.e., cooking procedures) with detailed information about precise quantities, shapes and manners, in a structurally organized format. In the next subsection, we examine how actions are expressed, through the analysis of types and forms of verbs.

4.3 Verb types and forms
4.3.1 Transitivity
In Cookbook, an overwhelming majority of verbs are lexically transitive or used transitively (90.9%, compared with 37.2% in News). When a lexical transitive verb is not available, a causative suffix, *(s)ase*, is used (e.g., *futtoo-sase-ru* 'make it boil'), making a generic agent the (unexpressed) subject. Intransitive verbs appear within subordinate clauses or noun-modifying clauses (e.g., *yawarakaku nat-tara* 'when (it)

becomes soft,' *futtoo shi-ta yu* 'boiling water') and provide conditional and other types of adverbial information for when and how actions should be performed. The pervasiveness of transitive verbs, which is linked to the high frequency of the object particle *o*, underscores the procedural nature of Cookbook.

4.3.2 Finite forms: plain vs. polite

At the sentence final position, the choice between the plain and polite forms is mostly a matter of convention. One of the five cookbooks consistently uses the polite form (*masu*) in every recipe. The remaining four cookbooks use the plain form consistently. However, three recipes from three different cookbooks with the predominantly plain style also contain some polite *masu* forms. They appear, not in step-by-step instructions, but either in the post-instruction comment section or the prefatory comment above the dish name. At least in some cookbooks, the choice between the plain and polite forms seems to be functionally motivated, and this is related to explicit awareness of the existence of the addressor and addressee. The use of plain form in the procedural steps keeps the text impersonal and detaches it from both the addressor and the addressee. The use of the polite form in the comment section, on the other hand, foregrounds the awareness of their existence and presents the text as social communication between the two. The association between the polite ending and the high degree of awareness of the addressee is recognized in previous studies. For example, the polite form is used when speakers ask questions, give personal comments/opinions during a story telling activity (Clancy 1982) or when authors write as if addressing the reader directly (Maynard 1991), provided that the speakers/writers are not personally close to the addressees. The functional and structural division between the plain and polite forms is also seen in other technical manuals; the plain form is used to describe procedural steps, while the polite form is used to provide further explanations and comments.[8]

4.3.3 Medial forms: *te* vs. *i/e*

Sentences in Cookbook are rather short (averaging 17.9 words per sentence, compared with 28.6 words per sentence in News), but regularly contain multiple clauses (an average of three clauses per sentence). Verbal clauses in the medial position frequently end in the continuative *te* (39.2%) or *i/e* (i.e., *renyoo-kei*) (45.8%). The choice between two clause-linking forms largely depends on the relative degree of continuity or integration between the two situations described in the *te* or *i/e* attached clause and the immediately subsequent clause within the local context. Following Ono (1990), continuity is measured in regard to: (1) the number of new participants in the second clause (0–4), (2) temporal continuity between the two clauses (simultaneous, immediate, lapse), (3) change of place (no change or change), and (4) whether or not a comma punctuation (、) is used after the first clause.

In the next example, (1), the *te*-attached clause (*zaru ni hiroge-te* 'spread on the strainer') has no new participant in the subsequent clause (*mizuke o kir-i* 'drain water'), has a simultaneous temporal relationship with the subsequent clause, has no change of place, and no comma. The *i/e*-attached clause

(*mizuke o kir-i*), in contrast, has two new participants in the subsequent clause (*shio, koshou o suru* 'put salt and pepper'), has an immediate temporal relationship with the subsequent clause, no change of place, and a comma.

>   (1) *te* and *i/e* medial forms in Cookbook
>   *zaru ni <u>hiroge-te</u> mizuke o <u>kir-i</u>, shio, koshou o suru.*
>   'Spread (<u>*hiroge-te*</u>) (the oysters) on the sieve to drain (<u>*kir-i*</u>) water, and put salt and pepper (on them).'

Overall, the *te* form indicates a higher degree of integration with the immediately subsequent clause than the *i/e* form, especially with respect to the number of newly mentioned participants (only 25.2% of *te*-clauses has one or more new participants in the subsequent clause, whereas the number is 67.1% with *i/e*-clauses), and temporal continuity to a lesser extent (21% of the *te*-clauses have a simultaneous temporal relationship with the subsequent clause while none of the *i/e*-clauses do). It is striking that the comma is used 99.4% of the time after *i/e*-clauses and only 22.4% after *te*-clauses. This seems to indicate that the degree of continuity is context-dependent. The choice between *te* and [*i/e* + comma] is not automatic based on any inherent characteristics of verbs or processes being described; it is a decision (however conscious or unconscious) of the writer (and/or the editor) about how to present the processes in each local context.

In News, there is a strong preference for using the *i/e* form (*i/e* 47.1%, *te* 12% of all medial forms), whether clauses represent sequential events, cause and effect, reason and situation, or simple listings of information. The *te* form is used only in the following situations: describing simultaneous actions/events, when the *i/e* form of a verb is one-syllable (e.g., *miru* 'see' → *mi*), when the first clause expresses the means through which the event/action in the second clause is accomplished, or when the *te* form is part of an idiomatic phrase (e.g., *to-shi-te* 'for the reason that...,' *to mi-te* 'positing that...'). Nonetheless, the comma appears with the *i/e* form much more frequently (94.4%) than with the *te* form (28%).

4.4 Introduction and tracking of major ingredients
In this subsection, we analyze how major ingredients in Cookbook are introduced and tracked, focusing on the arguments of predicates within main clauses. Major ingredients are introduced explicitly in a list as self-standing nouns or noun phrases. When they are introduced as noun phrases, subsequent mentions of the same referents in the step-by-step procedure become reduced in form. (e.g., *buta-usugiri-niku* 'thinly sliced pork' → *buta-niku* 'pork,' *momen-doufu* 'firm tofu' → *toufu* 'tofu'). The use of reduced forms is seen in News and other registers; hence it is not a special feature of Cookbook.

11

When ingredients require preparation, such as washing, peeling, and cutting, before they can be combined with other ingredients, they are marked with the topic particle *wa* in the second mention after they are introduced in the list of ingredients (e.g., *ninjin wa kawa o muite-oku*. 'As for the carrot, peel the skin.') This pattern is pervasive and consistent in cooking recipes (Aoyama 1987, Moriya 1993). After the preparation stage, subsequent mentions of the same ingredients are not marked with *wa*. Example (2) tracks the initial and subsequent mentions of tofu, a major ingredient in a dished named *niku-doufu* 'beef tofu.'

(2) Introduction and tracking of a major ingredient
1st: *momen-doufu* 'firm tofu' (List of ingredients)
2nd: *toufu wa*... 'As for the tofu...' (Step 1)
3rd: *toufu o kuwae*... 'add the tofu...' (Step 3)
4th: *5～6 pun Ø* (combined state) *niru* 'simmer Ø for 5–6 minutes' (Step 3)
5th: *2～3 pun Ø* (combined sate) *nite*... 'simmer Ø for 2–3 minutes...' (Step 3)

The second mention of tofu is marked by *wa* and the rest of the sentence following *wa* instructs an unexpressed agent to cut the tofu into bite-size pieces. The third mention of tofu is marked by the object marker *o*. Ellipsis is used for the fourth and fifth mentions, in which it refers not to the tofu alone but to a combined state of tofu and beef (4th mention) or of all the ingredients (5th mention). The use of ellipsis to refer to a combined state is regularly observed towards the final steps in Cookbook. Ellipsis is also observed when making a second or third reference to the same ingredient within the same step. The way ellipsis is used in Cookbook avoids unnecessary redundancy and ambiguous referent. In News, coherent structure and high topic persistence allow much more frequent use of ellipsis. Sometimes, in Cookbook, the step number (or the step number + the particle *no* + noun) is used to refer to the result of a previous step, that is, an intermediate state of the ingredient/dish (see Yamakata et al. 2017).

The particle *wa* is generally considered to mark topical (or thematic) information (e.g., Maynard 1981, 1987, Noda 1996). In a written narrative, for example, a topic (or theme) is the main character from whose point of view the story is told (Maynard 1981). In Cookbook, the particle *wa* does not mark the topicality of the referent, but functions as "a cohesive device between locally contrasting elements" (Clancy & Downing 1987: 49), where the contrasting elements are various ingredients that make up a dish. In the recipe for 'meat tofu,' for instance, tofu is not the only major ingredient; *gyuuniku* 'beef,' *wakegi* 'scallions,' and *enokidake* 'enoki mushrooms' are each marked with *wa* in the preparation stage (Step 1). While the cohesive function of the particle *wa* only works at the local discourse level in oral narratives (Clancy & Downing 1987), in Cookbook, it also contributes to the global text organization, as

the use of *wa* helps to distinguish the preparation stage from the rest of the processes. This is particularly helpful for readers when the preparation stage is dispersed across several numbered steps.

4.5 Text structure

All of the cookbook texts in the current dataset share the following structural features: (1) the dish name, (2) a photo of the finished product, (3) a list of ingredients under the heading *zairyoo* 'ingredients,' and (4) step-by-step instructions under the heading *tsukurikata* 'directions' (lit. 'how to make'). Figure 2 shows a typical layout of a cookbook recipe. The name of the dish is written on top in a font larger than the rest of the text. The list of ingredients is accompanied by information about quantities, amount, and size. Some minor ingredients such as condiments are visually grouped together with brackets, parentheses, or indentation, sometimes headed by an alphabetical letter (e.g., A) or composite name (e.g., dressing). The procedural instructions are always given in numbered steps. Each numbered step usually contains only one sentence (75.3%), sometimes two (19.3%), but rarely three or more. There are usually three to six steps in a recipe.

[Name of the dish]

[Photo of the dish]

[List of ingredients]   [Numbered steps]
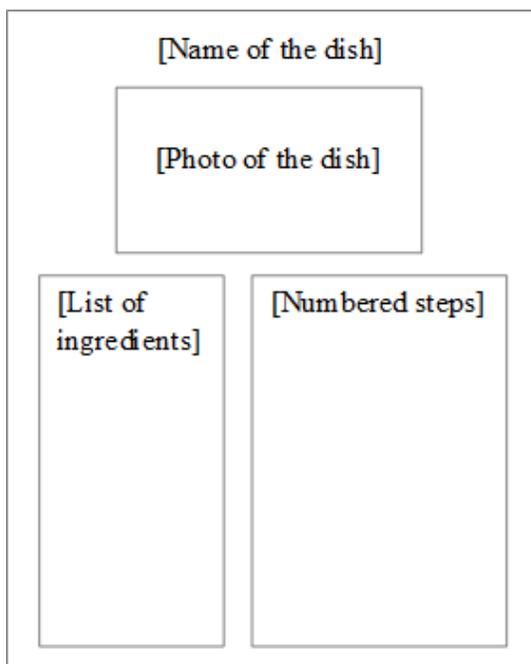
Figure 2. Schematic representation of the sections of cookbook recipes

The dish name provides a quick reference to essential information about the dish that is described in the main part of the text. The photo of the finished dish also functions to this end. Separation of the ingredient list and step-by-step instructions is conventionalized but also contributes to the visual clarity of

13

the text. The spatial separation and headings make it clear where to find information about ingredients and the procedural steps in one glance. This is important because readers of the recipes may have intermittent attention if they are looking at a recipe while cooking. The numbered steps are also conventional features of Japanese cookbook recipes (see Martinec 2003 and Strauss et al. 2018 for comparison of Japanese recipes with other languages). They also contribute to the visual clarity of the sequential actions. The use of numbers, instead of words such as temporal conjunctions, eliminate the need to follow the sentences carefully, and they make it less likely for the readers to lose their place while glancing at the steps.

4.6 Characteristics of cookbook recipes: summary

As mentioned in 4.1, the situational characteristics that distinguish cookbook texts from other planned and edited written texts such as news reports are the narrow topic of food/cooking and the major communicative purpose of providing food preparation instructions. In (3), the linguistic and structural features of Cookbook described in the previous subsections are linked to the two situational characteristics. The communicative purpose is further divided into procedurally driven aspects and clarity driven aspects.

(3) Associations between linguistic/structural features and situational characteristics

Narrow topic: food/cooking
- o Low(er) lexical diversity
- o High frequency of cooking terminology
- o Use of some specialized words

Communicative purpose: food preparation instructions

Procedurally driven aspects
- o High(er) proportion of verbs (as well as adverbs)
- o Specific and limited functions of particles such as *ni*
- o High frequency of quantity specifiers
- o Pervasiveness of transitive verbs (and the object particle *o*)
- o Systematic use of *te* and [*i/e* + comma] in the medial position

Clarity driven aspects
- o High frequency of text organizers
- o Bipartite structure (list of ingredients and step-by-step instruction)
- o Name of the dish and the photo of the finished product
- o Numbered steps with short sentences

- Use of the particle *wa* for preparation steps
- Use of step numbers to refer to the intermediate state of ingredients
- Consistent use of plain form in the procedural steps

While the topical features and the procedurally driven aspects of the communicative purpose are rather intuitive, to appreciate the clarity driven aspects, we need to take into account the assumed setting of communication in which some readers have more or less divided attention as they glance at the instructions while cooking. All features listed under the clarity driven aspects facilitate quick and easy processing of information. Three features, the bipartite structure, the contrastive use of the particle *wa*, and the anaphoric use of step numbers, are not found in other types of technical manuals (see Endnote 8).

5. Online commercial and user-generated recipes

Comparison of linguistic and organizational features of Cookbook with those of online commercial recipes (henceforth, Commercial) and online user-generated recipes (henceforth, User-generated) reveals that the two online recipe sub-registers combine elements of traditional recipes, to varying extents, and features related to a larger online context of information and social networks. In the following analyses, we included all features produced by recipe writers/editors and site creators, but excluded readers' comments, if any.

5.1 Situational characteristics

The three recipe sub-registers share the topic of food and cooking, the main purpose of providing food preparation instructions, and the type of addressees (unremunerated general public), but as shown in Table 3, they differ in other aspects. Cookbook and Commercial share the type of addressors, the relationship between addressors and addressees, and the production circumstances, while Commercial and User-generated share the medium and some interactiveness between addressors and addressees. At the same time, the type of interaction differs between Commercial and User-generated: indirect in the former and direct in the latter. Besides the main objective of offering cooking instructions, Commercial also has the purpose of providing an online network of related information, while User-generated has the purpose of creating an online social community among its users. The following subsections (5.2–5.5) demonstrate how these situational commonalties and differences are reflected in the linguistic features found in the three sub-registers.

15

Table 3. Situational characteristics of three sub-registers

	Cookbook	**Commercial**	**User-generated**
Topic	food/cooking	food/cooking	food/cooking
Addressee	general public	general public	general public
Addressor	expert	expert	amateur
Relationship	professional-learner	professional-learner	community members
Production circumstances	planned, revised, edited	planned, revised, edited	unedited/self-edited (individual variation)
Medium	printed	online	online
Interactiveness	none	some (indirect)	some (direct)
Purpose	provide cooking instructions	provide cooking instructions and a network of related information	provide cooking instructions and a social community

5.2 Vocabulary

5.2.1 Lexical diversity and categories

The additional purposes of the two online sub-registers besides providing cooking instructions are reflected in the overall higher number of words in these two sub-registers, given the same number of texts (Table 4; numbers also presented in Table 1). There are twice as many words in Commercial and over three times more words in User-generated than in Cookbook.

Table 4. No. of texts and words in the three sub-registers

	No. of text	Total No. of words	Average No. of words per text (range)
Cookbook	30	5,175	173 (89–540)
Commercial	30	11,503	383 (179–640)
User-generated	30	18,714	624 (331–1,002)

According to the Maas' index scores (Table 5), User-generated has the highest lexical diversity, followed by Commercial. The differences are statistically significant based on one-way ANOVA test, $F(2,87) = 75.32$, $p < .05$, $\eta_p^2 = .63$. Some linguistic features in Commercial and User-generated perform functions aimed at achieving objectives other than providing cooking instructions, such as involving readers as active participants and providing a network of related information (see Table 3 in 5.1).

Table 5. Maas index scores in the three sub-registers

	Maas (M)	SD
Cookbook	10.32	5.38

| Commercial | 22.01 | 8.42 |
| User-generated | 38.00 | 11.44 |

Turning to the distribution of lexical categories, we find a major division between Cookbook/Commercial and User-generated. Specifically, the former sub-registers contain more content words (nouns, verbs,[9] adjectives, and adverbs) than the latter. Figure 3 shows the frequency of content words and function words (particles, auxiliaries, prefixes, adnouns, conjunctions, interjections, and symbols) in each sub-register.
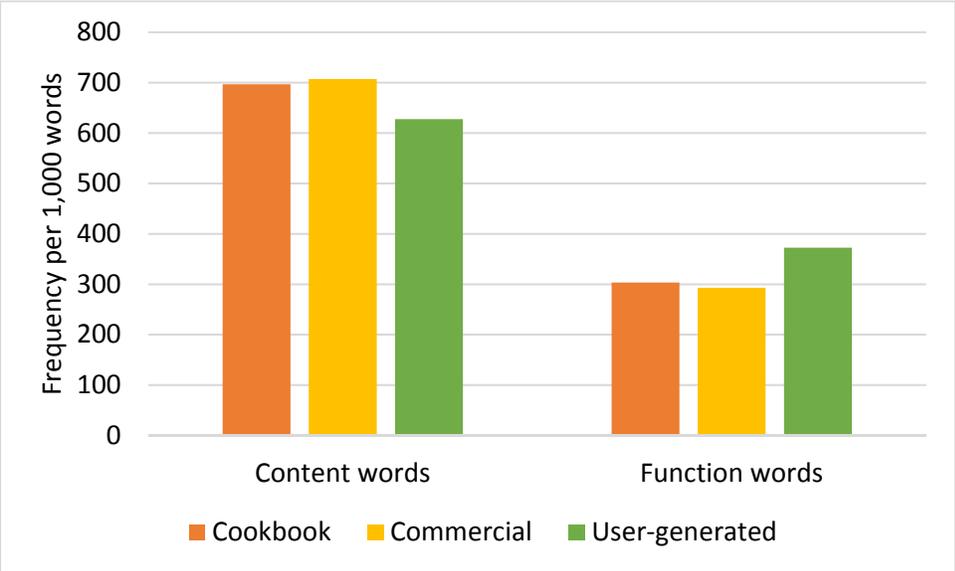

Figure 3. Frequency of content words and function words in the three sub-registers (per 1,000 words)

In particular, User-generated has the highest portion of symbols (1.4 times more than Commercial and 2 times more than Cookbook), auxiliaries (2.7 times more than Commercial and 3.6 times more than Cookbook), and interjections (Commercial and Cookbook do not contain interjections). While most symbols in Cookbook are text organizers and quantity specifiers, symbols in User-generated also work as affective markers. Affective markers express "a mood, attitude, feeling, and disposition, as well as degrees of emotional intensity vis-à-vis some focus of concern" (Ochs 1996: 410). For example, an exclamation mark or heart symbol following *arigatou* 'Thank you' may express a heightened emotion of gratitude. In User-generated, close to half of all symbols are affective (47.8%), followed by text organizers (30.8%) and quantity specifiers (6.8%). The higher proportion of auxiliaries in User-generated can be explained by the existence of interactional components. Close to half of the auxiliaries are

politeness markers (45.97%), followed by past tense markers (29.23%). In Commercial, past tense markers (45.35%) outnumber politeness markers (20.12%).

Interjections only appear in User-generated. 63.89% are lexical interjections and 36.11% are non-lexical interjections. Among the lexical interjections, a majority are *arigatou* 'Thank you' (82.07%), but there are also other expressions such as *konnichiwa* 'Hello' (6.52%) and *sumimasen/gomennasai* 'I'm sorry' (3.26%). The non-lexical interjections are called *kaomoji* 'emoticons' (lit. 'face characters'). There are many varieties, and forms can be quite elaborate. Emotions expressed with *kaomoji* include happiness, (*´∀`)♪ ＼(＾o＾)／, crying for joy or gratitude, (ToT) (;∀;), excitement, (＞▽＜) (*≧ з ≦), modesty and/or embarrassment, (*´ω`*) (*ˆˆ*), friendliness, (^o^) (o^▽^o)／, self-reproach, (_ _), and light-heartedness, (^_-)-☆, among others. Similar to physical facial expressions and gestures, these emotions are not mutually exclusive. They seem to be used to express heightened or subtle emotions that may be difficult to express with words alone or with simpler symbols such as the exclamation symbol for excitement. Whether or not to use *kaomoji* may be a stylistic choice, considering that 13/30 posters (43.33%) do not use *kaomoji* at all. Among *kaomoji* users, these elaborate emoticons are used purposefully and are closely linked to the interactional nature of User-generated. Over half of the *kaomoji* are used within the posters' replies to readers' comments while 30% are used within the numbered procedural steps (also see Section 6).

5.2.2 Frequent words, keywords, and specialized words

In the following list of most frequent words (Table 6), half of the words are shared among all three sub-registers (particles, *te*-form, quantities 1 and 2, verb 'do'). Words that are only shared between Cookbook and Commercial are quantity specifiers (fraction symbol, *f(b)un* 'minutes; for,' *oosaji* 'tablespoonful') and text organizers (bullet points). These words indicate the higher prominence of cooking instructions as the main content of these texts relative to User-generated. There are two words shared between Commercial and User-generated: *reshipi* 'recipe' and the past tense suffix *ta*. Common collocations of these two words indicate that the online recipe texts present their recipes as part of a larger context of recipe sites as archives of numerous interlinked recipes on the one hand and as a community for the site users on the other. The noun 'recipe' tends to co-occur with words such as *[name] san no* 'Mr./Ms. [name]'s,' *kono/kochira no* 'this,' and *mai* 'my' on the left and with *ichiran* 'catalog,' *o* (object particle), and *ID* 'ID' on the right. These phrases function to locate a particular recipe among other recipes. The past tense suffix *ta* co-occurs with words such as *tsukut* 'make,' *shi* 'do,' and *mashi* (polite suffix) on the left and *komento* 'comment,' *yo* (interactional particle), and *hito/yuuzaa* 'person/user' on the right. These clusters (e.g., *tsukut-ta komento* '"(I) made (it)" comment,' *sutanpu shi-ta yuuzaa* 'users who clicked ("I made it" sticker)') demonstrate that the recipe readers/users are incorporated as active participants in these sub-registers. Frequent words that only appear in User-generated can be categorized into politeness

markers (*masu*, *mashi*, *o-*, *desu*) and interactional markers (!, ♪, *ne*). They together point to the two-way interactiveness of User-generated.

Table 6. Most frequent words in the three sub-registers

	Cookbook		Commercial		User-generated	
	Word	Meaning/function	Word	Meaning/function	Word	Meaning/function
1	*o*	Particle	*o*	Particle	*o*	Particle
2	*ni*	Particle	*ni*	Particle	*no*	Particle
3	*te*	*Te*-form	*no*	Particle	*te*	*Te*-form
4	*no*	Particle	*te*	*Te*-form	*ni*	Particle
5	1	Quantity	1	Quantity/serving size	*ta*	Past tense
6	*wa*	Particle	*wa*	Particle	*reshipi*	Noun 'recipe'
7	2	Quantity	*de*	Particle	*de*	Particle
8	*oosaji*	Noun 'tablespoonful'	2	Quantity/serving size	*shi*	Verb 'do'
9	*de*	Particle	*ta*	Past tense	!	Exclamation symbol
10	4	Quantity/serving size	・	Bullet point	1	Quantity/number
11	*to*	Particle	／	Fraction symbol	*masu*	Polite suffix
12	*shi*	Verb 'do'	*f(b)un*	Noun 'minute; for'	*ga*	Particle
13	●	Bullet point	*reshipi*	Noun 'recipe'	♪	Music note symbol
14	*f(b)un*	Noun 'minute; for'	*suru*	Verb 'do'	*o*	Polite prefix
15	2	Procedural step	*to*	Particle	*desu*	Polite copula
16	*shio*	Noun 'salt'	g	Noun 'gram'	2	Quantity/number
17	1	Procedural step	*ga*	Particle	*mashi*	Polite suffix
18	*ga*	Particle	*shi*	Verb 'do'	*wa*	Particle
19	／	Fraction symbol	*oosaji*	Noun 'tablespoonful'	*ne*	Interactional particle
20	*ire*	Verb 'put in'	3	Quantity	*to*	Particle/conditional /quote

A comparison of keyword lists between Commercial and User-generated gives further insight into the differences between the two online sub-registers (Table 7).[10] In Commercial, all the keywords, except *ryoori*, *tori*, *san*, and *ta*, are part of the preformatted template of the website page. Even the exclamation

symbol, !, is part of existing phrases, such as *ii ne!* 'Like!,' *shea shiyoo!* 'Let's share!' and *minna no oishii ! komento* 'Everyone's delicious! comments.' One major purpose of the existing features is to encourage participation from the readers; some elements (*yaku* and *tat* as part of *yaku ni tat-ta* '(it) was useful') are tied to clickable buttons, and other elements (*ichiran*, *tooroku*, *miru*) direct viewers to other pages via hyperlinks. Another characteristic of Commercial is seen in the keywords *ryoori*, *san*, and the colon. These keywords are related to introducing and publicizing the cooking experts who are the creators of the recipes. Thus, Commercial also serves as a mediator between professionals and non-expert readers. This is different from how cooking experts are portrayed in Cookbook; they are represented as the (formal) writers of the texts.

In User-generated, about half of the keywords are existing features (*reshipi*, *ken*, *yo*, :, *stanpu*, *miru*, *repooto*, *kono*, and *ii*) while one-third are found in posters' comments (!, *desu*, *mashi*, *gozai*, *arigatou*, and ♡). The rest of the keywords have multiple usages. The fact that politeness markers (*desu*, *mashi*, *gozai*) and interactional features (!, *arigatou*, ♡) are a part of posters' comments demonstrates that these texts are directly interactive between writers/posters and readers/viewers rather than being mediated by site organizers, as in Commercial.

Table 7. Top 20 keywords in Commercial and User-generated

	Commercial		User-generated	
1	*reshipi*	Noun 'recipe'	*reshipi*	Noun 'recipe'
2	*komento*	Noun 'comment'	!	Exclamation symbol
3	!	Exclamation symbol	*ta*	Past tense
4	*ryoori*	Noun 'cooking'	♪	Music note symbol
5	*miru*	Verb 'see'	*desu*	Polite copula
6	*kono*	Adnoun 'this'	*mashi*	Polite suffix
7	*yaku*	Noun (part of idiom)	*ken*	Counter
8	*tat*	Verb (part of idiom)	*gozai*	Polite suffix
9	0	Number	*yo*	Interactional particle
10	*ta*	Past tense	*o*	Polite prefix
11	*ichiran*	Noun 'list/table'	*arigatoo*	Expression 'thank you'
12	*seisaku*	Noun 'creation'	♡	Heart symbol
13	*tookoo*	Noun 'contribution'	:	Colon (ID; date; hour)
14	*tsukut*	Verb 'make'	*sutanpu*	Noun 'sticker'
15	*tori*	Noun 'chicken	*miru*	Verb 'see'
16	*tooroku*	Noun 'register'	*ne*	Interactional particle

17	*san*	Suffix 'Mr./Ms.'	☆	Star symbol
18	*pointo*	Noun 'point'	*repooto*	Noun 'report'
19	*nichi/hi/ni*	Noun 'date'	*kono*	Adnoun 'this'
20	:	Colon (credits list)	*ii*	Adjective 'good'

Most of the specialized words found in Cookbook are also found in Commercial and User-generated. ([shape]-*giri*, verb compounds *maze-awaseru* and *mawashi-ireru,* and *kitsune-iro*). There are additional specialized words in Commercial and User-generated: *kuwae-mazeru* 'add and mix,' *hito-nitachi-saseru* 'bring to a gentle boil,' and *mori-tsukeru* 'serve/arrange (on a plate/bowl).' One term only used in User-generated is *tsuku-repo* or *repo* for short (both in hiragana) 'cooking report (lit. made-report).' In addition, the expression *dekiagari* or *kansei* 'completed' is found at the end of cooking procedures in a majority of User-generated (21/30) but not in Cookbook or Commercial. These expressions demonstrate the writers' (posters') higher involvement in the described processes, taking the first-person perspective of an assumed cook and user of the recipe. They also function as the coda (Labov & Waletzky 1967) of the cooking steps (the recipe proper). In narratives, a coda functions to bring the reader/listener back to reality (time of speaking). In User-generated, the coda *dekiagari* and *kansei* work to overtly mark the completion of cooking steps embedded in a larger social context. Such a function is not necessary in Cookbook or Commercial, as cooking steps comprise the major context together with the list of ingredients in these sub-registers.[11]

5.3 Verb types and forms

5.3.1 Transitivity

The proportion of transitive verbs, as compared with intransitive verbs, in Commercial (87.3%) and User-generated (81.6%) is slightly lower than that in Cookbook (90.9%). In addition, in Commercial and User-generated, intransitive verbs appear sentence-finally (recall that they only appear within subordinate and relative clauses in Cookbook). Sentence-final intransitive verbs appear within descriptions of dishes/recipes as well as information about cooking experts in the case of Commercial and messages to readers in the case of User-generated. Thus, in Commercial and User-generated, cooking instructions are placed within a larger context of various information related to the dishes/recipes and social interaction with the readers.

5.3.2 Finite forms: plain vs. polite

In Section 4.3.2, we saw that the choice between the plain and polite forms is largely a matter of convention in Cookbook, but also that in three individual texts there is a functional and structural division between the plain form, used for cooking instructions, and the polite form, used for prefatory and post-

instruction comments. This distributional pattern is more consistently observed in Commercial, as all texts in this sub-register contain pre- and post-instruction comments. Although there is a difference in the regularity of comments between Cookbook and Commercial, both sub-registers show consistency in the use of the plain and polite forms. In contrast, some discrepancies are observed in User-generated. 80% (24/30) of the posters use only the plain form in cooking instructions and the polite form in post-instruction comments, tips, and recipe backgrounds. The rest of the posters use the polite form throughout (3/30), use it in cooking instructions and backgrounds and the plain (casual) form in comments and tips (2/30), or use a mixture of polite and plain forms in prefatory and cooking instructions, the polite form in comments and tips, and the plain form in the background section (1/30). Despite some individual differences, a majority (80%) of posters follow the pattern of employing the plain form for technical aspects and the polite form for social communication (see Section 4.3.2).

A complication arises, in the sense of competing motivations, when posters in User-generated write replies to individual readers who contribute cooking reports and comments on the posters' recipes/dishes. Since the posters do not have a personal relationship with the readers, the use of the polite form would be the norm. However, using the polite form also evokes distance and formality. Some posters choose to use casual forms (6/30, 20%), but a majority (22/30, 73.3%) use the polite form. Among those who use the polite form in the replies, 86.4% (19/22) also use features that represent a (face-to-face) conversational style, such as an elongation symbol (〜), interactional particles (e.g., *ne*, *yo*, *yone*), emphatic tones (e.g., *tottemo* 'veeery,' *suggoi* 'reeeally,' *takkusan* 'maaany') (see Shinzato 2018), a laughter symbol ((笑) or w), and reflexive vocalization (e.g., *waa~*, *oo~*), as well as *kaomoji* 'emoticons' (lit. face characters). While the conversational features are also used by those who write exclusively in a casual style, *kaomoji* are only used by those who write in the polite form. This suggests that *kaomoji* are used for a specific function. The use of *kaomoji* allows posters to express friendliness to readers without also indicating personal closeness with them, as if they were friends or family members.

5.3.3 Medial forms: *te*, *i/e*, and *tara*

We saw in Section 4.3.3 that verbal clauses in the medial position predominantly take *i/e* or *te* forms in Cookbook. As shown in Figure 4, Commercial displays a very similar pattern as Cookbook. In these two sub-registers, procedural actions are connected with *i/e*-forms or *te*-forms while temporal conditions for these actions are expressed with the conditional *tara* 'when/if,' the temporal *made* 'until,' and *nagara* 'while.' [Intransitive verb + *tara*] focuses readers' attention on the completion of the target's change. Japanese has other conditionals and adverbial particles, such as *kara*, *toki*, and *to*, which could be used in place of *tara*, but the focus would be on the order of the actions/events as a whole (Moriya 1993: 119–120, 128).

Compared with Cookbook and Commercial, the proportion of *tara* is much higher, and the proportion of *i/e* and *te* forms is much lower in User-generated. This is because posters describe temporal and optional conditions and intermediate states of ingredients more frequently, and *tara* is also used for linking actions, like *te* and *i/e* forms. This use of *tara* for sequential actions is called clause chaining and is reported as a characteristic of spoken discourse (Ono & Iwasaki 2002, Ono & Jones 2009). In the present data, 9/20 (45%) texts include this use of *tara* (a total of 11 examples). Another difference between User-generated and Cookbook/Commercial is that the former contains a larger number of *te* forms than *i/e* forms. This is due to individual differences in the use of *te* and *i/e* forms in User-generated. 11/30 texts (36.7%) show a clear preference for one form over the other (using one form 4 times or more than the other), in some cases exclusively using one form (4/30, 13.3%). Some of these texts use the comma (、) to show higher discontinuity with either *i/e* or *te* form and use the conditional *tara* for linking actions, as mentioned earlier.[12]
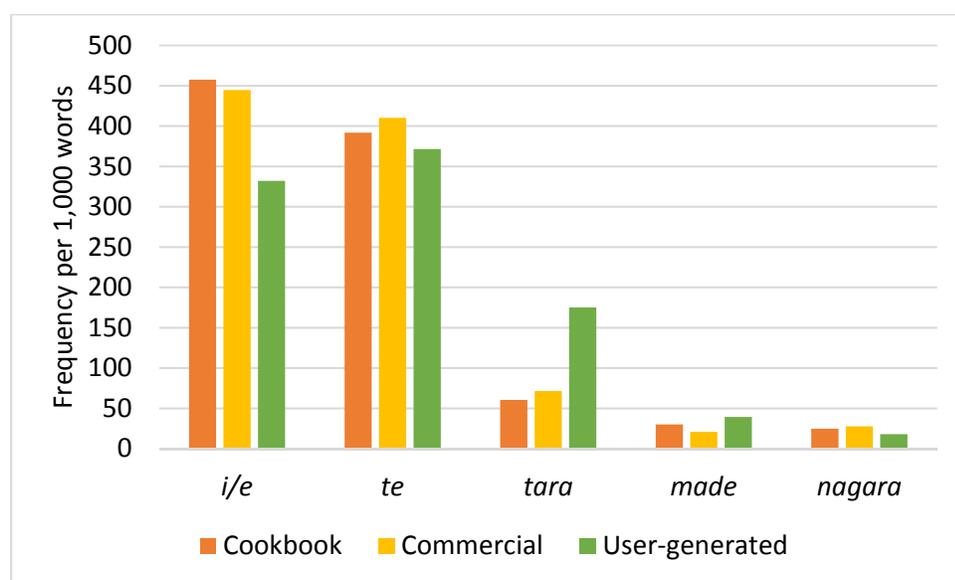

Figure 4. Frequency of medial verb forms in the three sub-registers (per 1,000 words)

The differences in frequency and usage patterns of medial forms between Cookbook/Commercial and User-generated indicate higher conformity in professionally-edited texts. Unedited or self-edited texts show a much wider variety. A choice between *i/e* and *te* forms based on the relative degree of continuity does not seem to be a part of at least some native speakers' linguistic (production) knowledge. Some use other resources such as the clause chaining *tara* and a punctuation mark to accomplish functions that *i/e* and *te* forms mainly perform in professionally-edited recipe texts. A question remains, however, as to

whether individual amateur writers of User-generated consulted other recipe texts, including professionally-edited ones, in writing their own.

5.4 Introduction and tracking of major ingredients

The pattern of major ingredient introduction and subsequent references found in Cookbook is also found in Commercial. An ingredient is introduced in the list of ingredients as a bare noun or noun phrase, referred to in a reduced or full form, with the particle *wa*, in preparation steps, and without *wa* in the subsequent non-preparation steps. Ellipsis is also used when a second or third reference is made to the same ingredient within the same step or towards the final steps when a reference is made to a combined state of some or all ingredients.

User-generated texts show more variations. The use of *wa* in the preparation stage for at least one ingredient is only found in half of the texts (16/30). One-third of the texts simplified the preparation stage by (1) using a relative clause when introducing major ingredients in the procedural steps or (2) referring to a group of major ingredients (e.g., all the vegetables are grouped together and referred to as ☆). In one case, the preparation steps are completely omitted, referring to a group of ingredients assumed to have gone through some preparation already. The bypassing of some steps indicates that posters assume a certain level of shared background knowledge with readers. Although posters do not have a personal relationship with readers, through the latter's cooking reports and comments, they become informed about readers' knowledge state, preference, and so forth. In fact, in many of the texts, posters add post-instruction comments and tips based on reader feedback. The less explicit, simplified characteristic of User-generated resembles community recipes (Cotter 1997) in which writers assume shared knowledge with their community members. Five of the User-generated recipes also used the demonstrative pronoun *soko* 'there' or *sore* 'that' to refer to the intermediate state of some ingredient(s) from the previous step. These deictic demonstratives are also less explicit ways of referencing than the use of nouns or step numbers.

5.5 Text structure

The structural features of Cookbook described in Section 4.5 are shared among all three sub-registers, but the two online sub-registers also show some additional structural features. The following lists present text structural features that are shared among all three sub-registers (4), features shared by the two online sub-registers (5), and features only found in Commercial (6) or in User-generated (7). Figures 5 and 6 show the typical layout of Commercial (Figure 5) and User-generated recipes (Figure 6).

(4) Text features found in all three sub-registers

24

- o Names of dishes
- o Photos of finished products
- o Lists of ingredients and quantities under the heading *zairyoo* 'ingredients'
- o Step-by-step procedural instructions

(5) Text features found in the online sub-registers
- o Print/save buttons
- o Social networking service share links (Facebook, Twitter)
- o Categories/classification of recipes with hyperlinks
- o Hyperlinks to other related recipes
- o Clickable opinion buttons
- o Pre-instruction comments
- o Post-instruction comments/tips
- o Recipe users' reports/comments, including date, photo, comment, icon, and screen name (handle) of each user

(6) Text features only found in Commercial
- o Name of instructors (cooking expert)
- o Photos and bios of instructors
- o Hyperlink to instructors pages
- o Hyperlink to instructors' other recipes

(7) Text features only found in User-generated
- o Screen names and icon of posters
- o Recipe IDs and posting dates
- o Recipe background information (called *oitachi* 'personal history' or *kikkake* 'initial motivation')
- o Email/send to cell phone buttons
- o Posters' replies to users' reports/comments
- o Messages/announcements within the numbered cooking steps (20/30)

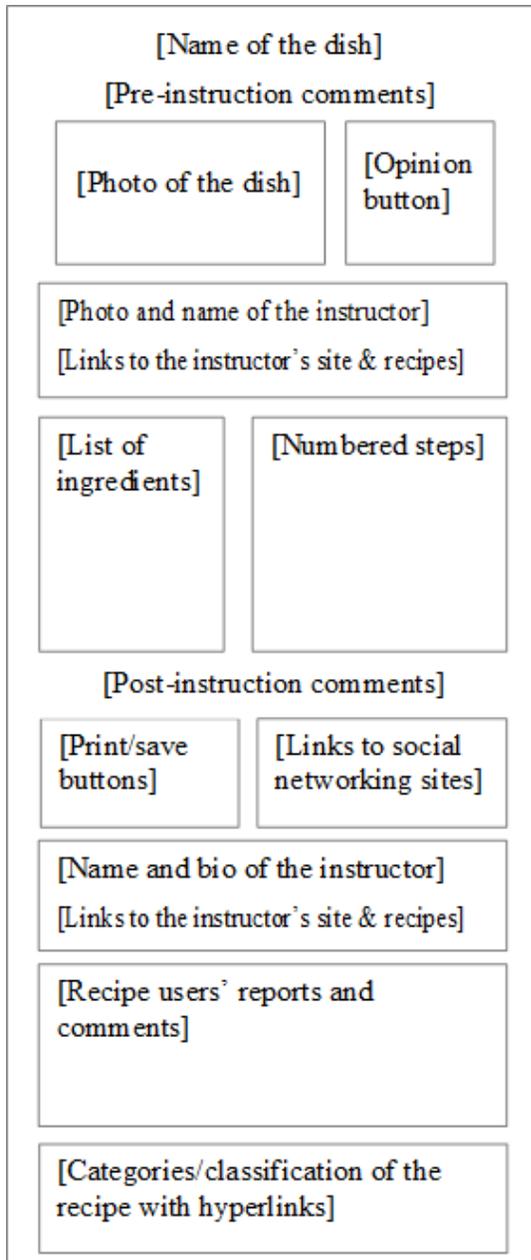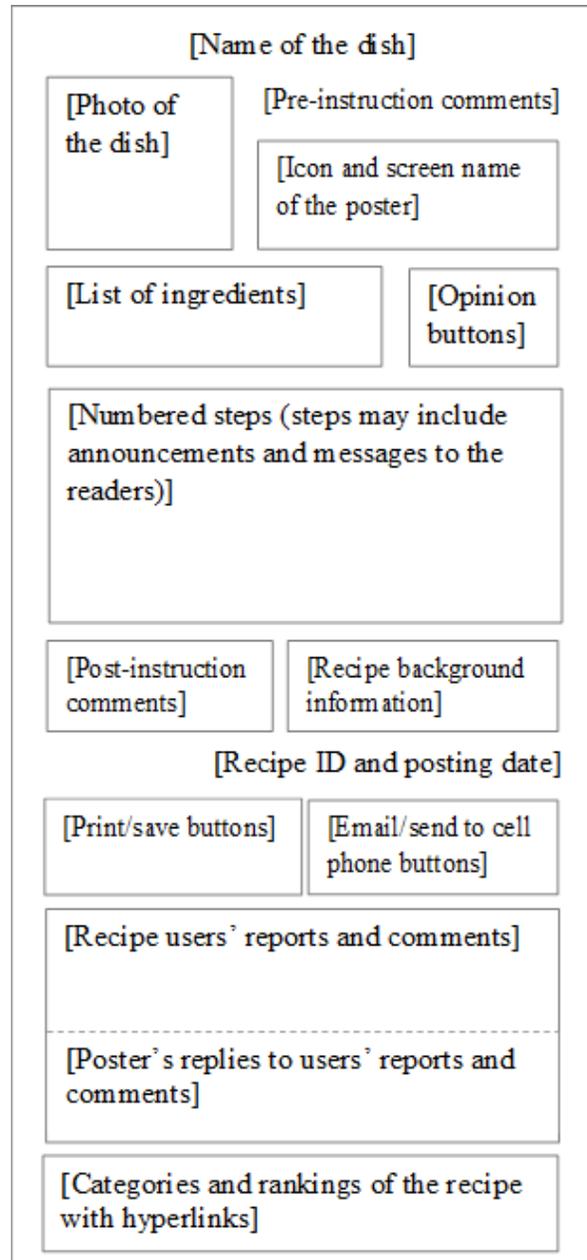| [Name of the dish]	[Name of the dish]
[Pre-instruction comments]	[Photo of the dish] [Pre-instruction comments]
[Photo of the dish] [Opinion button]	[Icon and screen name of the poster]
[Photo and name of the instructor] [Links to the instructor's site & recipes]	[List of ingredients] [Opinion buttons]
[List of ingredients] [Numbered steps]	[Numbered steps (steps may include announcements and messages to the readers)]
[Post-instruction comments]	[Post-instruction comments] [Recipe background information]
[Print/save buttons] [Links to social networking sites]	[Recipe ID and posting date]
[Name and bio of the instructor] [Links to the instructor's site & recipes]	[Print/save buttons] [Email/send to cell phone buttons]
[Recipe users' reports and comments]	[Recipe users' reports and comments]
	[Poster's replies to users' reports and comments]
[Categories/classification of the recipe with hyperlinks]	[Categories and rankings of the recipe with hyperlinks]

Figure 5. Schematic representation of Commercial   Figure 6. Schematic representation of User-generated

The structural features shared by all three sub-registers, (4), contribute to the visual clarity of the text, providing readers with quick access to essential information. Features shared between Commercial and User-generated, shown in (5), provide readers with choices to save, print, share the current recipe, explore other recipes, and participate by contributing cooking reports or simply clicking buttons such as *oishii!* 'delicious!,' and *ii ne!* 'like!'

What separates Commercial and User-generated is how interactional components are integrated. In Commercial, interaction only happens indirectly between some readers (cooking report contributors) as they sometimes refer to each other's reports. These reports are written for other recipe readers/users. The commercial recipe sites are designed as a learning and feedback-sharing space where non-professional readers learn how to make dishes from professionals who are presented as instructors. In User-generated, on the other hand, direct (though asynchronous) interaction takes place between recipe posters and recipe readers/users. Cooking reports are addressed to posters instead of other readers, as the writers of the reports expect to receive replies from the posters to whom they report. The sites are designed as a community space for recipe posters and recipe seekers. The posters who share their home recipes and background narratives present themselves not as experts but as community members and equals to the recipe readers/seekers. It is striking that 20/30 texts (66.7%) contain the posters' comments and messages not related to cooking procedures within the numbered steps. These comments include announcements about the recipe rankings and recognitions, announcements about reaching a certain number of cooking reports, and thank you messages to the recipe users. An informal internet survey of some posters suggests that they genuinely enjoy sharing their recipes and receiving comments that indicate that others had successful experiences with their recipes. Some posters also comment that they enjoy cooking at home more and they are better at cooking because of their posting experience. Many also say that they initially began posting their recipes for their own record; the community and interactional aspect thus seems to evolve over time as more recipes and comments are shared.

6. Similarities and differences among the three recipe sub-registers

All three sub-registers share features that embody the narrow topic of food as well as the common purpose of providing easy-to-follow food preparation instructions for readers who may have dispersed attention. The centrality of this purpose and related elements, however, differ among the three. Cookbook shows the highest homogeneity and convergence with respect to the two said dimensions. User-generated exhibits the most deviation and most variations. Commercial seems to be somewhere between Cookbook and User-generated.

To understand the observed differences, we need to consider three additional situational factors: production circumstances and editorial control, medium, and relations among participants. As shown in (8) below, Cookbook and Commercial share the property of high editorial control. Commercial and User-generated have two aspects in common: the medium of an online network and reader participation. The online medium leads to the role of the sub-registers as archives of numerous recipes. The ways readers are involved, however, are not the same. In Commercial, readers can express their opinions and give feedback, through given means, in one-way communication. In User-generated, the communication is

two-way as posters can respond and reply to the readers' feedback and comments. Some posters even alter the function of numbered steps and include their messages to the readers. Thus, there is a three-way distinction in the communication pattern of the three sub-registers: single unidirectional in Cookbook (from writers to readers), dual unidirectional in Commercial (from writers to readers, and from readers to other readers), and bidirectional in User-generated (between writers and readers).

User-generated differs from the other two sub-registers in two ways: it has low or no editorial control, and it involves direct interaction between writers and readers. Two items that are listed under both of these situations, the use of conversational language/*kaomoji*, and posters' messages and announcements within numbered steps, suggest that interactional features emerge in the least controlled environment. It is also notable that morphological and grammatical features (medial and finite verb morphology and the use of the particle *wa*) are affected by the level of editorial control. The fact that a high degree of variation is seen in the low control environment implies that these features are not crucial in accomplishing the central purpose of recipes, but they may give a sense of polished and professional writing.

(8) Three (additional) dimensions of recipes

Production circumstances: Editorial control

High control (Cookbook/Commercial)

- o Systematic use of *te* and [*i/e* + comma]
- o Systematic use of plain and polite forms
- o Consistent use of the particle *wa* for preparation steps

Low control (User-generated)

- o Variations in the use of *te*, *i/e*, and *tara*
- o Variations in the use of plain and polite forms
- o Variations in describing (including simplifying or omitting) preparation steps
- o Use of conversational features and *kaomoji*
- o Writers' messages and announcements within numbered steps

Medium

Online network (Commercial/User-generated)

- o Numerous hyperlinks
- o Use of recipe identifiers/locators

Relations among participants

Readers' participation (Commercial/User-generated)

- o Clickable opinion buttons
- o Template for cooking reports and comments

Direct interaction between writers and readers (User-generated)

- o High frequency of politeness markers
- o Use of interjections
- o Use of conversational features and *kaomoji*
- o Template for replies to readers' comments
- o Writers' messages and announcements within numbered steps
- o Use of *dekiagari* or *kansei* 'completed' at the end of cooking procedures

7. Conclusion

In this paper, we described the regularity and variation in three Japanese recipe sub-registers. All three sub-registers include linguistic and textual features that are linked to the topic of food and cooking and the purpose of providing food preparation instructions that can be easily followed and quickly understood. At the same time, only the sub-registers with high editorial control show a high degree of consistency and conformity in the grammatical features that contribute to the perspicuous cooking instructions. Unanswered questions include whether or not amateur writers of user-generated recipes refer to professionally edited recipes in writing their own, and if so, how such a practice affects their language use. While providing clear cooking instructions is the sole purpose of cookbook recipes, online commercial and online user-generated recipes serve other communicative purposes. Both sub-registers serve as a repository and online reference center for other recipes and related information. The online medium also allows both sub-registers to involve readers as active participants via the contribution of comments and cooking reports. In online commercial recipes, communication is one-way from some readers to other readers. In online user-generated recipes, communication is two-way between writers and readers. In addition, low editorial control allows direct interaction between the two parties, resulting in patterns and features of communication that are not found in the other two sub-registers. These findings suggest that a single register may encompass a range of communicative goals and other situational factors, which shape, and are sometimes shaped by, the patterns of language use within particular contexts.

Endnotes

---

[1] On one of the two user-generated recipe sites, there is a category called *pro no reshipi* (pro's recipes), but one needs to pay monthly fees to access these recipes, and we did not use any recipes from this category.

[2] All recipes included in the current dataset are written for adult Japanese speakers. The choice of specific sources is partially based on availability. Specific data sources are listed below.

Cookbook: (1) Egami, K. (2002) *Hajimete no ryoori*. Nihon-bungeisha, (2) *Kantan okazu 698-shu*. (1997) Shufu-to-seikatsusha, (3) Hata, K. & Kondo, K. (1996) E*igo de nihon-ryoori*. Kodansha, (4) Makuuchi, H. (2001) *Kodomo reshipi*. Shufu-no-tomosha, (5) Maki, H. (2009) *Uchi gohan*. San-ei shobo.

Commercial: (1) *E-reshipi*, Retrieved from: https://erecipe.woman.excite.co.jp/, (2) *Kewpie 3-pun kukkingu*, Retrieved from: http://hicbc.com/tv/kewpie/, (3) *Kyoo no ryoori*, Retrieved from: https://www.kyounoryouri.jp/

User-generated: (1) *Cookpad*, Retrieved from: https://cookpad.com/, (2) *Rakuten reshipi*, Retrieved from: https://recipe.rakuten.co.jp/

[3] As pointed out by one of the reviewers, a comparison of two registers may exaggerate situational and linguistic characteristics that happen to differ between the two registers. We believe a comparative perspective still leads to a more comprehensive understanding of a register than an individual analysis of a single register.

[4] In order to get a representative sample of news reports, we sampled news articles from six different categories from four major news sources in Japan. Specific data sources are as follows: *Asahi Shimbun*, www.asahi.com; *Mainichi Shimbun*, mainichi.jp; *Sankei Shimbun*, www.sankei.com; *Yomiuri Shimbun*, www.yomiuri.co.jp. The news reports come from the following categories: society (*shakai*), politics (*seiji*), business/economy (*keizai*), international affairs (*kokusai*), culture (*bunka*), and sports.

[5] POS categories used by IPADIC are as follows: *meishi* 'noun,' *joshi* 'particle,' *doushi* 'verb,' *jodoushi* 'auxiliary,' *fukushi* 'adverb,' *keiyoushi* 'adjective,' *settougo* 'prefix,' *setsuzokugo* 'conjunction,' *kandoushi* 'interjection,' and *kigou* 'symbol.' See Asahara and Matsumoto (2003) for more information about the categories.

[6] Readers' attention is not included in Biber and Conrad's discussion of situational characteristics, but it would be categorized under the setting, which refers to "the physical context of the communication" (2009: 44).

[7] We used the online search site: http://www.kotonoha.gr.jp/shonagon/

[8] In our informal examination of ten manuals for different products (a watch, TV, car, rice cooker, cooking stove, cell phone, washer, external drive, vacuum, and baby carriage), eight manuals (80%) followed this pattern. Two manuals, one for an external drive and the other for a vacuum, used the polite form for both steps and comments. This may be due to relatively simple procedures involved in using these products.

[9] The proportion of verbs is higher in all three recipe sub-registers than in News (11.28%).

[10] Keyword lists for Commercial and User-generated are obtained by comparing each corpus against Cookbook corpus. We used the log-likelihood as the keyness test.

[11] It is interesting to note that the expression *dekiagari* is often used in children's how-to books and picture books that deal with cooking; in these environments, *dekiagari* functions more like a traditional narrative coda, such as *oshimai* 'the end' and *medetashi-medetashi* 'happily ever after' in folktales and children's stories.

[12] The use of commas after the *i/e* form is much more consistent in Cookbook (99.4%) and Commercial (97.4%) than in User-generated (62.4%).

References

Akiyama, M. (2002). Syoku o meguru hyoogen: Resipi no bunshoo/buntai ni tsuite [Expressions concerning food: Texts and styles of recipes]. *Kokugo Kenkyuu 65*, 36–46.

Anthony, L. (2014). AntConc (Version 3.4.4w) [Software]. Tokyo, Japan: Waseda University. Available from: www.laurenceanthony.net/software.

Aoyama, F. (1987). Ryoori no bunshoo ni okeru teidaika no yakuwari [The role of topicalization in recipe texts]. In *Keiryoo kokugogaku to Nihongo shori: riron to ooyoo* (pp. 258–303). Tokyo: Akiyama Shoten.

Asahara, M. & Matsumoto, Y. (2003) IPADIC version 27.0 Users Manual. Nara Institute of Science and Technology. Available from: http://chasen.naist.jp/snapshot/ipadic/ipadic/doc/ipadic-ja.pdf.

Barlow, M. & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford: CSLI.

Basso, K. H. (1974). The ethnography of writing. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 425–432). Cambridge: Cambridge University Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1994). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 31–56). Oxford: Oxford University Press.

Biber, D. (1995). *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, D., Davies, M., Jones, J. K., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: a multi-dimensional analysis. *Corpora 1*, 7–38.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language 82*(4), 711–733.

Cotter, C. (1997). Claiming a piece of pie: How the language of recipes defines community. In A. Bower (Ed.), *Recipes for reading: Community cookbooks, stories, histories* (pp. 51–72). Amherst: University of Massachusetts Press.

Clancy, P. (1982). Written and spoken style in Japanese narratives. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 55–76). Norwood: Ablex.

Clancy, P. & Downing, P. (1987). The use of *wa* as a cohesion marker in Japanese oral narratives. In J. Hinds, S. Maynard & S. Iwasaki (Eds.), *Perspectives on topicalization: The case of Japanese wa* (pp. 3–56). Amsterdam: Benjamins.

Crystal, D. & Davy, D. (1969). *Investigating English style*. London: Longman.

Diemer, S. (2013). Recipes and food discourse in English – a historical menu. In C. Gerhardt, M. Frobenius & S. Ley (Eds.), *Culinary linguistics: The chef's special* (pp. 139–156). Amsterdam: Benjamins.

Du Bois, J. (1987). The discourse basis of ergativity. *Language* 63, 805–855.

Ferguson, C. (1994). Dialect, register, and genre: Working assumptions about conventionalization. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 15–29). Oxford: Oxford University Press,

Fischer, K. (2013). The addressee in the recipe: How Julia Child gets to join you in the kitchen. In C. Gerhardt, M. Frobenius & S. Ley (Eds.), *Culinary linguistics: The chef's special* (pp. 103–117). Amsterdam: Benjamins.

Halliday, M.A.K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.

Hopper, P. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–175). Mahwah, NJ: Lawrence Erlbaum.

Hymes, D. (1974). *Foundations in sociolinguistics: an ethnographic approach*. Philadelphia: University of Pennsylvania Press.

Kittredge, R. (1982). Variation and homogeneity of sublanguages. In R. Kittredge & J. Lehrberger (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 107–137). Berlin: de Gruyter.

Kim, Y-J. & Biber, D. (1994). A corpus-based analysis of register variation in Korean. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 157–181). New York: Oxford University Press.

Kudo, T. (2013). MeCab (Version 0.996). Available from: taku910.github.io/mecab.

Labov, W. & Waletzky, J. (1967). Narrative analysis. Oral version of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle: University of Washington Press.

Martinec, R. (2003). The social semiotics of text and image in Japanese and English software manuals and other procedures. *Social Semiotics 13*(1), 43–69.

Maynard, S. (1981). The given/new distinction and the analysis of the Japanese particles -*wa* and -*ga*. *Papers in Linguistics 14*, 109–30.

Maynard, S. (1987). Thematization as a staging device in the Japanese narrative. In J. Hinds, S. Maynard, & S. Iwasaki (Eds.), *Perspectives on topicalization: The case of Japanese wa* (pp. 57–82). Amsterdam: John Benjamins.

Maynard, S. (1991). Pragmatics of discourse modality: A case of *da* and *desu/masu* forms in Japanese. *Journal of Pragmatics 15*, 551–582.

Moriya, M. (1993). Tetsuzukiteki chishiki no bunshoo − ryooribun − no shisasuru mono [Implications of procedural knowledge texts – recipes]. *Kooza Nihongo Kyooiku 28*, 114–130.

Naganuma, M. (2006). Honyaku ni okeru 'meishika' to iu bunpooteki hiyu [Grammatical metaphor of nominalization in translation]. *Interpretation Studies 6*, 15–28.

Noda, H. (1996). *Wa to ga* [*Wa* and *ga*]. Tokyo: Kuroshio Publishers.

Ochs, E. (1996). Linguistic resources for socializing humanity. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 407–437). New York: Cambridge University Press.

Ono, T. (1990). *Te*, *i*, and *ru* clauses in Japanese recipes: A quantitative study. *Studies in Language 14*(1), 73–92.

Ono, T. & Iwasaki S. (2002). Toward an understanding of 'sentence' in spoken Japanese discourse: Clause-combining and online mechanisms. In K. Kataoka & S. Ide (Eds.), *Culture, interaction, and language* (pp. 103–131). Tokyo: Hitsuji shobo.

Ono, T. & Jones, K. (2009). Conversation and grammar: Approaching so-called conditionals in Japanese. In J. Mori &A. Ohta (Eds.), *Japanese applied linguistics* (pp. 21–51). New York: Continuum.

Shinzato, R. (2018). *Amari/anmari/anma* and *totemo/tottemo* in history and discourse: Interaction of negative polarity and positive polarity. In M. Hudson, Y. Matsumoto & J. Mori (Eds.), *Pragmatics of Japanese: Perspectives on grammar, interaction and culture* (pp. 49–73). Amsterdam: Benjamins.

Strauss, S., Chang, H. & Matsumoto, Y. (2018). Genre and the cultural realms of taste in Japanese, Korean, and U.S. online recipes. In M. Hudson, Y. Matsumoto & J. Mori (Eds.), *Pragmatics of Japanese: Perspectives on grammar, interaction and culture* (pp. 219–243). Amsterdam: Benjamins.

Yamakata, Y., Caroll, J. & Mori, S. (2017). A comparison of cooking recipe named entities between Japanese and English. In *Proceedings of the 9th workshop on multimedia for cooking and eating activities* (pp.7–12). Melbourne, Australia.

Zwicky, A. M. & Zwicky, A. D. (1982). Register as a dimension of linguistic variation. In R. Kittredge & J. Lehrberger. (Eds.), *Sublanguage: Studies of language in restricted semantic domains* (pp. 213–218). Berlin: Walter de Gruyter.