

Attractive Multimodal Instructions

Describing Easy and Engaging Recipe Blogs

Ielka van der Sluis, Jarred Kiewiet de Jonge
Center for Language and Cognition Groningen (CLCG)
University of Groningen The Netherlands
{i.f.van.der.sluis, j.j.b.kiewiet.de.jonge}@rug.nl

Abstract

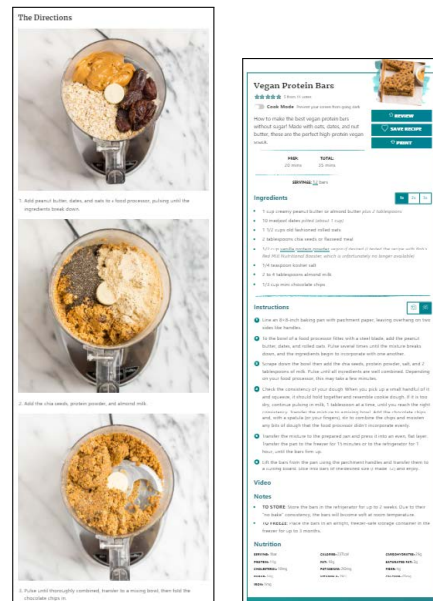
This paper presents a corpus study that extends and generalises an existing annotation model which integrates functional content descriptions delivered via text, pictures and interactive components. The model is used to describe a new corpus with 20 online vegan recipe blogs in terms of their Attractiveness for at least two types of readers: vegan readers and readers interested in a vegan lifestyle. Arguably, these readers value a blog that shows that the target dish is Easy to Make which can be inferred from the number of ingredients, procedural steps and visualised actions, according to an Easy to Read cooking instruction that displays a coherent use of verbal and visual modalities presenting processes and results of the cooking actions involved. Moreover, added value may be attributed to invitations to Engage with the blog content and functionality through which information about the recipe, the author, diet and nutrition can be accessed. Thus, the corpus study merges generalisable annotations of verbal, visual and interaction phenomena to capture the Attractiveness of online vegan recipe blogs to inform reader and user studies and ultimately offer guidelines for authoring effective online multimodal instructions.

Keywords: multimodal instruction, document design, corpus analysis, vegan recipe blogs

1. Introduction

1.1. Multimodal Recipe Blogs

Recipes have been a source of inspiration for structured text analysis for some time now (Bieñ et al., 2020; DiMeo and Pennell, 2018; Floyd and Forster, 2017; Mori et al., 2012; Görlach, 1992). In addition, recipes are often composed of verbal and visual modes and thus allow for the evaluation of the effectiveness of multimodal presentations for a variety of readers as well as users in multiple respects (e.g., attractiveness, comprehension, performance). Recipe blogs are a specific type of online documents that share recipes, cooking tips and food-related content. A recipe blog presents a procedural instruction that guides users through the steps involved to prepare a dish (Van der Sluis and Mellema, Submitted). Figure 1 illustrates that recipe blogs present the instruction in two formats (Bowker, 2021; Domingo et al., 2014). The blog as a whole presents an Instruction with Pictures (IWP), a



(a) IWP of MI 3.

(b) RC of MI 3.

Figure 1: Source: <https://www.wellplated.com/vegan-protein-bars/>.

multimodal step-by-step instruction combining text and pictures, and allows for additional dynamic content (e.g., adds, videos). At the end of the blog a Recipe Card (RC) is offered, which presents all the necessary steps and ingredients to prepare the dish in text.

With the growing popularity of mindful dieting, a large body of blogs offer recipes for crafting nutritious and healthy dishes at home (Guha and Gupta, 2020), but what makes a blog attractive? The study presented in this paper examines the means that authors of online content use to attract their public. Retrieval and analysis of blogs are interesting because authors simultaneously employ a range of semiotic elements (e.g., text, pictures, videos, interactive features). To explore the blog authors' use of available modes and functions to attract potential online readers and users, we conducted a small and focused corpus study. Based on existing approaches and findings in multimodal and online content analysis, the corpus study is offered as a starting point to conduct future reader and user evaluations and to support the further development and automation of our preliminary notion of Attractiveness.

The corpus solely contains recipes for vegan nutrition bars ie. compact and portable snacks typically crafted from plant-based ingredients like nuts, seeds, fruits, and grains that serve as a source of essential nutrients catering to health-conscious and environmentally-aware consumers. Studying vegan blogs is timely because the past decade displays a noticeable shift and steady increase in the adoption of an exclusively plant-based lifestyle (Kustar and Patino-Echeverri, 2021; Kamiński et al., 2020; Schösler et al., 2012). In the cooking domain, this trend is mirrored in recipe blogs that support a vegan diet (Asano and Biermann, 2019). Given the abundance of recipe blogs and the variation in which the food preparation procedures in them are presented it is of interest to identify the characteristics that make a blog attractive to both vegans as well as those that are merely interested in a vegan lifestyle. Recipes for vegan nutrition bars in particular offer potential to convince blog users due to their convenience and popularity as a healthy and

nutritious snack or meal replacement option that can contribute to supporting a healthier diet in a relatively quick and easy way (Bansal et al., 2022; Jovanov et al., 2021).

The corpus study was set up to answer the following research question: Which means do authors of step-by-step recipe blogs for vegan nutrition bars use to attract potential users? Sections 1.2 and 1.3 introduce the background for a notion of Attractiveness which is operationalised using three aspects: Easy to Make, Easy to Read and Engagement.

1.2. Attractive Vegan Recipe Blogs

Arguably, recipes are Easy to Make dependent on the number of ingredients involved, the number of procedural steps described and the availability of visual presentations of those steps. The recipe becomes Easy to Read when the instructional parts of the blog display coherence and consistency in terms of their text content (RC vs. IWP) and coherence in the text-picture combinations within the IWP (cf. Bowker, 2021; Li and Xie 2020; Kang, 2010). Engagement requires alignment of the author's values with the needs and preferences of the blog users (Cooper et al., 2022; Machnee, 2019) as well as useful and playful content (Mainolfi et al., 2022; Liao et al., 2013). Given the wealth of online food recipes, blog authors are compelled to grab the attention of blog users. Independent of the intrinsic qualities of a recipe, the presence and professionalism of the blog pictures is crucial in influencing readers to choose a recipe (Starke et al., 2021), although effects of visual content, colourfulness, appearance of human faces and text-picture relations depend on the social medium platform (Li and Xie, 2020). Apart from quality pictures, the presence and credibility of the blog author is important. Bloggers should be knowledgeable, influential, passionate, transparent and reliable (Kang, 2010; Rubin and Liddy, 2006). A blogger's appearance hinges on a learned, positive writing style while credibility, trust and authenticity are gained through sharing personal stories (Machnee, 2019). At last, users increasingly consider nutritional characteristics when

selecting recipes to support informed decisions that align with their health and dietary needs (Cooper et al., 2022; Cheng et al., 2021; Rokicki et al., 2018; Trattner et al., 2018; Elswailer et al., 2017; Van Pinxteren et al., 2011; Freyne and Berkovsky, 2010). Accordingly, in a recipe blog the information about nutrition and diet should be present and easy to find.

1.3. Annotating Multimodal Instructions

Multimodality requires interdisciplinary research because multiple modes cohere and make meaning together (Bateman et al., 2017; Jewitt, 2009). Multimodal recipes rely on a combination of textual directions and visual cues (Ganier, 2012, 2000; Mayer, 2005). An instructive text assists people in executing a task through a step-by-step description of procedural information, usually presented in a numbered list of actions (Karreman and Loorbach, 2013). Alongside the procedural information instructions also contain control information (Van der Sluis et al., 2022; Karreman et al., 2005), encompasses non-procedural supplementary details relevant to the described process such as warnings, explanations, conditions etc. In instructions these two types of information work in tandem to ensure that users have the necessary knowledge and understanding to complete a task successfully (Ummelen, 1997).

Bateman (2014) describes coherence relations between text and pictures in terms of how one mode expands the meaning of the other (cf. Van der Sluis and Mellema, Submitted; Halliday and Matthiessen, 2013; Kress and Van Leeuwen, 2001; Barthes, 1977). Elaboration occurs when information is restated in another mode at a similar level of generality. For instance, an action is described in the text as a process (e.g., mix ingredients) and the related picture presents the result of that action (e.g., the dough as a result from mixing the ingredients). Enhancement on the other hand involves providing qualifying information related to aspects such as time, place, manner, reason, purpose, and other circumstantial

restrictions. For instance, the text describes an action (e.g., stir a substance) and the picture shows that the action is performed using a particular utensil (e.g., a whisk is used to stir a substance).

In multimodal instructions procedures can be described in terms of the actions involved. Recently, human action annotation and retrieval gained interest in multiple domains, media and applications (Pustejovsky and Krishnaswamy, 2022; Alikhani et al., 2019; Pustejovsky, 2018; Van der Sluis et al., 2018; Pustejovsky et al., 2017; Zhang et al., 2016; Lev et al., 2016; Laptev et al., 2008). The annotation model proposed to describe the vegan nutrition bar recipe blog corpus employs and extends the action-based PAT annotation model (Van der Sluis et al., 2022, 2017, 2016b)¹. The PAT model has been used to describe (parts of) multimodal instructions according to the following steps:

1. The instructional text is split into clauses;
2. The clauses are identified as either Action clauses or Control Information clauses;
3. The text clauses and the accompanying instructional pictures are described using functional attributes (e.g., Action Type, Action Status, Action Aspect, Control Information, Specification);
4. Coherence relations are described as compositions of text and picture annotations.

The generalisability of the PAT model is shown by annotating multimodal instructions in different domains, such as first-aid instructions (Van der Sluis et al., 2017) and cooking instructions (Van der Sluis and Mellema, Submitted; Van der Sluis et al., 2016b), through the annotation of multiple document types e.g., illustrated texts; instructional videos (Vijfvinkel et al., 2018) and instructional comics (Wildfeuer et al., 2022). The current corpus study presents a further development which merges

¹In the Pictures And Text or PAT project (<https://www.rug.nl/let/pat>), the PAT workbench (Van der Sluis and Redeker, 2019; Van der Sluis et al., 2016a) was built as an online tool designed to systematically describe multimodal documents.

annotation of different phenomena i.e. text, pictures, text-picture relations and interaction components to achieve a description of a context dependent notion of Attractiveness while further exploring the model's generalisability by describing online multimodal instructions.

2. Method

2.1. Corpus

The online recipe blogs for vegan nutrition bars were collected according to the following selection criteria:

- the blog includes an IWP and a RC;
- the IWP text describes the cooking procedure in at least three steps;
- the IWP includes at least three pictures visualising different stages in the cooking procedure;
- the RC has at least three procedural steps.

The vegan nutrition bar corpus contains 20 online recipes that were derived from eight distinct sources to allow a comparison between recipes from the same website while also ensuring a diverse representation across multiple sources. The corpus consists of four distinct parts with 5 recipes each: Part 1 contains 5 blogs from 5 different websites: Eat with Clarity², Vegan Huggs³, Well Plated⁴, Hummusapien⁵, and Minimalist Baker⁶. Part 2, 3 and 4 contain 5 recipes respectively from Veggie World⁷, All-Purpose Veggies!⁸, Eating Bird Food!⁹.

2.2. Annotation Model

The corpus study was set up to answer the following research question: Which means do authors of step-by-step recipe blogs for vegan

nutrition bars use to attract potential users? Attractiveness is operationalised using three aspects: Easy to Make, Easy to Read and Engagement, where the description of the notions Easy to Make and Easy to Read applies to particular parts of the blog namely the Instruction with Pictures and the Recipe Card, while the description of Engagement applies to the blog as a whole. The annotation model was largely based on the findings discussed in Section 1 of this paper. The annotation model was crafted and applied by two annotators that improved their work through multiple rounds of discussions until they agreed on the resulting model and the corpus description.

2.2.1. Easy to Make and Easy to Read

Conceivably, food preparation becomes or appears easier when a recipe includes only a few ingredients, when the procedure includes only a few steps and when the steps are visualised (cf. Yajima and Kobayashi, 2009). The blogs are described accordingly, using a notion Easy to Make that includes: (1) the number of necessary ingredients; (2) the number of steps in which the procedure is presented in the IWP; and (3) the number of visualisations of the procedural steps presented in the text.

A recipe becomes Easier to Read when the presentation in the instructional parts of the recipe blog displays coherence in terms their text content as well as coherence in combining text and pictorial information (cf. Kang, 2010). An action-based approach was taken to describe the coherence of the Instruction with Pictures and the Recipe Card for each blog in the corpus. The models described by (Van der Sluis and Mellema, Submitted; Van der Sluis et al., 2016b) were used as a starting point. Table 1 presents the text, pictures, and text-picture relation categories. The text clauses are annotated as Action or Control Information (Van der Sluis et al., 2022) Action clauses and visualised actions in pictures are annotated in terms of Status (i.e. Obligatory, Alternative, Conditional) and Aspect (i.e. Process, Result), where the Aspect value in the pictures is dependent on whether any utensils are included

²<https://eatwithclarity.com/>

³<https://veganhuggs.com/>

⁴<https://www.wellplated.com/>

⁵<https://www.hummusapien.com/>

⁶<https://minimalistbaker.com/>

⁷<https://veggieworldrecipes.com/>

⁸<https://allpurposeveggies.com/>

⁹<https://www.eatingbirdfood.com/>

in the visualisation. The Control Information clauses include Warning, Condition, Manner, Advice, Explanation, Motivation, Purpose and Situation Sketch. The text-picture relations are described in terms of Layout (i.e. Index, Proximity) and Content (i.e. Enhancement, Elaboration). The content relations are described in terms of meaning expansions, given a particular action that is presented in the two modes (Bateman, 2014).

2.2.2. Engagement

Engagement is described in terms of the presence of the following Text, Picture and Interaction attributes in the blog as a whole.

The following Text attributes are described:

- Attention Grabber - introduction text that reels in the audience such as “These vegan protein bars are a cookie dough flavored treat you’re going to love” (MI R1).
- Author Welcome - explicit greeting from the authors e.g., “Hey there! We’re jasmine and chris” (MI R2).
- Diet Legend - keys that specify the diets for which the recipe is suitable (e.g., VG, V, DF for respectively Vegan, Vegetarian, Gluten free).
- Location Diet Legend - place in the blog where the Diet Legend is offered (Top, Bottom, NA).
- Nutrition Facts - alimentary types and quantities included in the recipe (i.e. fat, carbs, sugars, protein, vitamins and minerals).
- Location Nutrition Facts - place in the blog where the Nutrition Facts are offered (i.e. Top, Bottom).

Pictures are described as follows:

- Author Portrait - picture of the blogger.
- Teaser - picture of the end result.
- Ingredients - picture of prepped but uncooked ingredients.
- Recommendation - picture of other recipes.

Included Interaction aspects are:

- Jump to Recipe - button to go to the RC.
- Link to Author - pointer to blogger details.
- Social Handles - pointers to the blogger’s social media.

- Rate Option - evaluate the recipe on a scale.
- Comment Option - write recipe evaluation.
- Tick-off function - boxes to indicate that ingredients are handy.

3. Analysis

3.1. Easy to Make and Easy to Read

Table 2 presents an overall description of the four parts of the Vegan Nutrition Bar Corpus that indicate in how far the recipes are Easy to Make. The 5 recipes from Veggie World contain the most ingredients, steps and pictures compared to the other subsets in the corpus. The average number of steps and pictures are balanced within each of the corpus parts. The number of ingredients and the number of steps seem unrelated e.g., Part 1 and Part 4 include more ingredients than steps.

In terms of Easy to Read, Table 3 presents the frequencies and percentages of Action Status and Control Information in the IWPs and RCs. The corpus has 1020 clauses: 654 Action and 375 Control Information clauses. The RCs contain more clauses ($N = 580$) than the IWPs ($N = 440$), with similar distributions of Actions and Control Information within the IWPs and RCs (IWP $\approx 63\%$ versus RC $\approx 37\%$). Most Action clauses (IWP = 221; RC = 299) present Obligatory Actions. The most frequent Control Information clauses present the Manner in which to perform an action ($N = 78$) and the Purpose for carrying out an action ($N = 76$).

Table 4 presents the frequencies and percentages of Action and Control Information clauses in the four corpus parts. The number of Actions varies between the subsets with a maximum of 202 actions in Veggie World and a minimum of 105 in All-Purpose Veggies!. The sets do not vary much in the number of Control Information clauses ($N \approx 94$).

Table 5 presents the text-picture relations in the IWPs in terms of Action Status and Action Aspect per corpus part. The IWPs contain 147 visualised actions and 279 verbalised actions. All pictures present Obligatory Actions, while the texts also contain Alternative ($N = 27$) and Conditional Actions ($N = 31$). The actions in





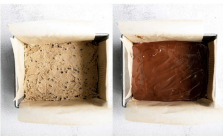


Text Attribute	Value	Description	Example (Source)
Action Status	Obligatory	An action that must be executed to perform the task successfully.	"Melt the dark chocolate chips in a tall glass." (MI 10)
	Alternative	An action that can be executed as a replacement of another action.	"(add more milk)...or water" (MI 14)
	Conditional	An action that can or must be executed under particular circumstances.	"then coat in melted chocolate." (MI 11)
Action Aspect	Process	The action is described as a process/in progress.	"Sprinkle with some flaky salt" (MI 18)
Control Information	Warning	The presentation addresses a possible danger.	"Be careful to avoid burning the coconut" (MI 20)
	Condition	The presentation specifies a condition or circumstance for an action to be performed.	"Once your coconut has cooled," (MI 20)
	Manner	The presentation addresses the way in which an action must be executed.	"until everything is evenly coated" (MI 19)
	Advice	The content of the presentation gives a recommendation on how to execute the action (not mandatory).	"I suggest storing these vegan protein bars in the fridge" (MI 18)
	Explanation	The presentation offers more information on how to execute the action.	"Each will give it a slightly different hue of green." (MI 6)
	Motivation	The presentation addresses a positive feeling or action.	"and enjoy!" (MI 6)
	Purpose	The presentation addresses the goal of executing the action.	"to encourage it to melt." (MI 5)
Situation Sketch	The content of the presentation displays a state in the procedure.	"Now it's time to make your filling." (MI 8)	
Picture Attribute	Value	Description	Example (Source)
Action Status	Obligatory	An action that must be executed to perform the task successfully.	 (MI R9)
Action Aspect	Process	The action is visualised with utensils and/or human hands.	 (MI R5)
	Result	The situation after completing an action, shown without utensils or hands.	 (MI R9)
Relation Attribute	Value	Description	Example (Source)
Layout	Index	Picture and text are related via the use of numbers, letters or titles.	 "4. Now add the mixture to the dates..." (MI R2)
	Proximity	Picture and text are related because they are positioned near to each other and integrated in the text. Reading direction is more important than physical distance on the page.	 (MI R1)
Content	Enhancement	Shows tools/hands to illustrate how the textualized action is performed.	 "Whisk together the oat flour, protein powder and salt." (MI R1)
	Elaboration	Provides additional information, in terms of, provisions of a result state in specific details without tools/hands present.	 "Add on top of the bars"(MI R1)

Table 1: Easy to Read attributes to describe Text, Pictures and Text-Picture relations.

the IWP text are always verbalised as a Process. Visualised actions appear as a Process showing utensils (N = 86) or as a Result (N = 61), showing the derived end state of an action. The differences between the corpus parts are substantial; Veggie World employs mostly

Process visualisations, while the other subsets display more variation in Action Aspect.

Table 6 presents the Layout and Content relations between the IWP text and pictures. Indices are not used much in the corpus, only the Vegan Huggs recipe in Part 1 includes enumer-

P	Source	Ingredients	Steps	Pics
1	5	8.2	5.4	4.2
2	1	11.2	11.4	11.6
3	1	5.4	6.4	6.0
4	1	7.6	6.2	6.6
All	1	8.7	7.9	6.8

Table 2: Easy to Make - Number of Sources and averages for Ingredients, instructional Steps and instructional Pictures per corpus Part and in the whole corpus.

ation to relate the text and pictures. Elaboration relations between text and pictures (N = 90), where the pictures present the result of a particular action are most frequent. Enhancement relations appear in 55 cases and mostly in the Veggy World blogs (N = 29). Two pictures are not related to a clause, 15 clauses are related to more than one picture (N = 35) and 899 clauses have no relation to any picture.

3.2. Engagement

Table 7 presents the frequencies and percentages for the Text, Picture and Interaction attributes to describe how authors invite user Engagement. In Text all blogs include Attention Grabbers and Nutrition Facts. In 2 of 20 blogs authors do not include an explicit welcome greeting. In 9 of 20 blogs the Diet Legend is omitted, which means that there is no indication about the suitability of the recipe for consumers with particular dietary constraints. The 9 Diet Legends that are included are always offered at the Top of the blog, while the Nutrition Facts are always offered at the bottom of the blog close to or as part of the RC.

The Picture attributes display that all blogs show a Teaser at the top of the blog that exemplifies the envisioned vegan nutrition bars. All blogs except one include a picture of the blog author. All blogs except one include Recommendations to other vegan recipes. Only 9 of 20 blogs include an image of the ingredients necessary to prepare vegan nutrition bars.

The Interaction attributes display that links to the Recipe card at the Bottom of the blog are included in all blogs. Also links to more

information about the author and the social media pages of the author are usually present. The means for a scaled or written evaluation of the recipes are also always included. The Tick-off function appears only in 9 of 20 blogs.

4. Discussion and Conclusion

The corpus study outlined in this paper offers a starting point to integrate different phenomena in online content with which the Attractiveness of online multimodal instructions that employ multiple modes (i.e. text, pictures, interaction components) can be described. The description provides a context dependent view on 20 vegan nutrition bar recipes from 8 sources constituted in three notions that offer insight in whether the blogs are Easy to Make, Easy to Read and Engagingly presented. In this case study the Attractiveness aspects were operationalised on the basis of existing findings from studies on multimodal communication, online content and the food domain, the newly developed categories may be complemented and improved in future work. For instance, in terms of Easy to make aspects such as preparation and cooking times or the availability of ingredients are likely of importance. In terms of Engagement, currently not all the attributes have equivalents in the described modalities. For example, the list of ingredients that is usually offered in the RC text was not included in the annotation model, while a picture of the ingredients was. Similarly, the nutritional facts are solely described as Text, while conceivably nutritional facts may also be visualised (cf. packaging of food products). Further grounding of the categories in terms of cultural and societal preferences are in order. For instance, the effectiveness of Process and/or Result visualisations in combination with verbalised Process actions needs further evaluation in a context of use, perhaps differentiating between novice and expert cooks. In the blog domain evaluation of the merit and/or annoyance of adds, videos and other dynamic content and of functions such as ticking off ingredients seems valuable. Finally, an extended description of author presence and credibility could be informed by

Attribute	Value	IWP		RC		Total	
		N	%	N	%	N	%
Action Clauses	Obligatory	221	79.2%	299	51.6%	520	51.0%
	Alternative	27	9.7%	33	5.7%	60	5.9%
	Conditional	31	11.1%	34	5.9%	65	6.4%
Total		279	63.4%	366	63.1%	654	63.2%
CI Clauses	Manner	29	6.6%	49	8.4%	78	7.6%
	Purpose	35	8.0%	41	7.6%	76	7.5%
	Condition	25	5.7%	32	5.5%	57	5.6%
	Advice	24	5.5%	33	5.7%	57	5.6%
	Warning	18	4.1%	20	3.4%	38	3.7%
	Motivation	15	3.4%	17	2.9%	32	3.1%
	Explanation	12	2.7%	17	2.9%	29	2.8%
Situation Sketch	3	0.7%	5	0.9%	8	0.8%	
CI Total		161	36.6%	214	36.9%	375	36.8%
Clause Total		440	100%	580	100%	1020	100%

Table 3: Easy to Read - Frequencies and percentages of Action and Control Information clauses in IWPs and RCs.

	Part	IWP		RC		Total	
		N	%	N	%	N	%
Action Clauses	1	46	10.5%	117	20.2%	163	16.0%
	2	91	20.7%	111	19.1%	202	19.8%
	3	52	11.8%	53	9.1%	105	10.3%
	4	90	20.5%	85	14.7%	175	17.2%
	All	279	63.4%	366	63.1%	645	63.2%
CI Clauses	1	30	6.8%	72	12.4%	102	10.0%
	2	39	8.9%	56	9.7%	95	9.3%
	3	46	10.5%	39	6.7%	85	8.3%
	4	46	10.5%	47	8.1%	93	9.1%
	All	161	36.6%	214	36.9%	375	36.8%
		440	100%	580	100%	1020	100%

Table 4: Easy to Read - Frequencies and percentages of Action and Control Information clauses in IWPs and RCs per corpus Part.

profile factors like expertise, identity disclosure, trustworthiness, content quality and personal appeals (Rubin and Liddy, 2006).

Although the two annotators that conducted the study used various rounds in which the annotation model and the corpus description was discussed and improved, the effort needs further evaluation in terms of inter-annotator agreement to obtain an indication of the difficulty of the annotation task and to examine in which ways the model can be improved, complemented and made generalisable as to apply to other online instructive blog content. In addition, prompting large language models on

the classification and generation of attractive instructions could further strengthen the exploratory results offered in this paper. Large databases containing cooking instructions, as well as videos of people executing them (e.g., Yagcioglu et al., 2018; Carvalho et al., 2018; Salvador et al., 2017; Regneri et al., 2013; Rohrbach et al., 2012a; Rohrbach et al., 2012b) demonstrate that a combination of text-based models with visual information can significantly improve the understanding and assessment of action descriptions. Recent initiatives in natural language processing and generation are promising (e.g., Tu et al., 2022; Pustejovsky

Attribute	Value	Part	IWP Text		IWP Pictures	
			N	%	N	%
Action Status	Obligatory		221	79.2%	147	100%
	Alternative		27	9.7%	0	0%
	Conditional		31	11.1%	0	0%
AS Total			279	100%	147	100%
Action Aspect	Process	1	46	100%	12	46.2%
		2	91	100%	55	94.8%
		3	52	100%	6	20.0%
		4	90	100%	13	39.4%
		All	279	100%	86	58.5%
	Result	1	0	0.0%	14	53.8%
		2	0	0.0%	3	5.2%
		3	0	0.0%	24	80.0%
		4	0	0.0%	20	60.6%
		All	0	0.0%	61	41.5%
AA Total			279	100%	147	100%

Table 5: Easy to Read - Frequencies and percentages of Action Status and Aspect in IWP Text and Pictures.

Part	Layout				Content			
	Index		Proximity		Enhancement		Elaboration	
	N	%	N	%	N	%	N	%
1	7	4.8%	19	13.1%	11	7.6%	15	10.3%
2	0	0%	58	40.0%	29	20.0%	29	20.0%
3	0	0%	28	19.3%	6	4.1%	22	15.2%
4	0	0%	33	22.8%	9	6.2%	24	16.6%
All	7	4.8%	138	95.2%	55	37.9%	90	62.1%

Table 6: Easy to Read - Frequencies and percentages of Layout and Content relations per corpus part.

Category	Attribute	N	%
Text	Attention Grabber	20	100%
	Author Welcome	18	90%
	Diet Legend	11	55%
	Nutrition Facts	20	100%
Picture	Author Portrait	19	95%
	Teaser	20	100%
	Ingredients Pic	9	45%
	Recommendation	19	95%
Interaction	Jump to Recipe	20	100%
	Link to Author	19	95%
	Social Handles	20	100%
	Rate Option	20	100%
	Comment Option	20	100%
	Tick-off Function	9	45%

Table 7: Engagement - Frequencies and percentages for Text, Picture and Interaction attributes for the whole corpus.

et al., 2021). Thus, annotation of cooking instructions serves to build systems that understand and extract practical knowledge from written instructions, enabling them to offer guidance or to perform procedural tasks. However limitations of computational tools for automatically identifying and categorising actions in instructions (Van der Sluis et al., 2018, Zhang et al., 2012) still require human intervention as an essential guiding factor. We advocate reader and user studies to explore the relevance of annotation models, to inform further annotation efforts and to inform guidelines for authoring multimodal instructions.

5. Acknowledgements

We are grateful for the positive and constructive comments of our ISA reviewers.

6. Bibliographical References

- Malihe Alikhani, Sreyasi Nag Chowdhury, Gerard De Melo, and Matthew Stone. 2019. Cite: A corpus of image-text discourse relations. *arXiv preprint arXiv:1904.06286*.
- Yuki M Asano and Gesa Biermann. 2019. Rising adoption and retention of meat-free diets in online recipe data. *Nature Sustainability*, 2(7):621–627.
- Uma Bansal, Aastha Bhardwaj, Som Nath Singh, Sucheta Khubber, Nitya Sharma, and Vasudha Bansal. 2022. Effect of incorporating plant-based quercetin on physicochemical properties, consumer acceptability and sensory profiling of nutrition bars. *Functional Foods in Health and Disease*, 12(3):116–127.
- Roland Barthes. 1977. *Image-music-text*. Macmillan.
- John Bateman, Janina Wildfeuer, and Tuomo Hiippala. 2017. *Multimodality: Foundations, research and analysis—A problem-oriented introduction*. Walter de Gruyter GmbH & Co KG.
- John A Bateman. 2014. Multimodal coherence research and its applications. In *The pragmatics of discourse coherence*, pages 145–177. John Benjamins.
- Michał Bień, Michał Gilski, Martyna Maciejewska, Wojciech Taisner, Dawid Wisniewski, and Agnieszka Lawrynowicz. 2020. Recipenlg: A cooking recipes dataset for semi-structured text generation. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 22–28.
- Skyler Bowker. 2021. [How to write a recipe post](#).
- Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. 2018. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 35–44.
- Xiaolu Cheng, Shuo-Yu Lin, Kevin Wang, Y Alicia Hong, Xiaoquan Zhao, Dustin Gress, Janusz Wojtusiak, Lawrence J Cheskin, and Hong Xue. 2021. Healthfulness assessment of recipes shared on pinterest: natural language processing and content analysis. *Journal of Medical Internet Research*, 23(4):e25757.
- Kelly Cooper, Ozgur Dedehayir, Carla Riverola, Stephen Harrington, and Elizabeth Alpert. 2022. Exploring consumer perceptions of the value proposition embedded in vegan food products using text analytics. *Sustainability*, 14(4):2075.
- Michelle DiMeo and Sara Pennell. 2018. *Reading and writing recipe books, 1550–1800*. Manchester University Press.
- Myrrh Domingo, Gunther Kress, Rebecca O’Connell, Heather Elliott, Corinne Squire, Carey Jewitt, and Elisabetta Adami. 2014. Development of methodologies for researching online: The case of food blogs.
- David Elsweller, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, pages 575–584.
- Janet Floyd and Laurel Forster. 2017. *The Recipe Reader: Narratives-Contexts-Traditions*. Routledge.
- Jill Freyne and Shlomo Berkovsky. 2010. Recommending food: Reasoning on recipes and ingredients. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 381–386. Springer.
- Franck Ganier. 2000. Processing text and pictures in procedural instructions. *Information Design Journal*, 10(2):146–153.

- Franck Ganier. 2012. Cognitive models of processing procedural instructions. *Commun. Technol*, 10:39.
- Manfred Görlach. 1992. Text-types and language history: The cookery recipe. *History of Englishes: New methods and interpretations in historical linguistics*, pages 736–761.
- Kritika Bose Guha and Prakhar Gupta. 2020. Growing trend of veganism in metropolitan cities: Emphasis on baking. *PUSA Journal of Hospitality and Applied Sciences*, 6:22–31.
- Michael Alexander Kirkwood Halliday and Christian MIM Matthiessen. 2013. *Halliday's introduction to functional grammar*. Routledge.
- Carey Jewitt. 2009. *The Routledge handbook of multimodal analysis*, volume 1. Routledge London.
- Pavle Jovanov, Marijana Sakač, Mihaela Jurdana, Zala Jenko Pražnikar, Saša Kenig, Miroslav Hadnadev, Tadeja Jakus, Ana Petelin, Dubravka Škrobot, and Aleksandar Marić. 2021. High-protein bar as a meal replacement in elite sports nutrition: a pilot study. *Foods*, 10(11):2628.
- Mikołaj Kamiński, Karolina Skonieczna-Żydecka, Jan Krzysztof Nowak, and Ewa Stachowska. 2020. Global and local diet popularity rankings, their secular trends, and seasonal variation in google trends data. *Nutrition*, 79:110759.
- Minjeong Kang. 2010. Measuring social media credibility: A study on a measure of blog credibility. *Institute for Public Relations*, 4(4):59–68.
- Joyce Karreman and Nicole Loorbach. 2013. Use and effect of motivational elements in user instructions: What we do and don't know. In *IEEE International Professional Communication 2013 Conference*, pages 1–6. IEEE.
- Joyce Karreman, Nicole Ummelen, and Michaël Steehouder. 2005. Procedural and declarative information in user instructions: What we do and don't know about these information types. In *IPCC 2005. Proceedings. International Professional Communication Conference, 2005.*, pages 328–333. IEEE.
- Gunther R Kress and Theo Van Leeuwen. 2001. Multimodal discourse: The modes and media of contemporary communication. (*No Title*).
- Anna Kustar and Dalia Patino-Echeverri. 2021. A review of environmental life cycle assessments of diets: plant-based solutions are truly sustainable, even in the form of fast foods. *Sustainability*, 13(17):9926.
- Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. 2008. Learning realistic human actions from movies. In *2008 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE.
- Guy Lev, Gil Sadeh, Benjamin Klein, and Lior Wolf. 2016. Rnn fisher vectors for action recognition and image annotation. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pages 833–850. Springer.
- Yiyi Li and Ying Xie. 2020. Is a picture worth a thousand words? an empirical study of image content and social media engagement. *Journal of marketing research*, 57(1):1–19.
- Chechen Liao, Pui-Lai To, and Chuang-Chun Liu. 2013. A motivational model of blog usage. *Online Information Review*, 37(4):620–637.
- Leigh Machnee. 2019. Authority, credibility and trust in vegan blogs: Methods used by content creators in the presentation of information. Master's thesis, Department of Computer and Information Sciences, University of Strathclyde.
- Giada Mainolfi, Vittoria Marino, and Riccardo Resciniti. 2022. Not just food: Exploring the influence of food blog engagement on

- intention to taste and to visit. *British Food Journal*, 124(2):430–461.
- Richard E Mayer. 2005. *The Cambridge handbook of multimedia learning*. Cambridge university press.
- Shinsuke Mori, Tetsuro Sasada, Yoko Yamakata, and Koichiro Yoshino. 2012. A machine learning approach to recipe text processing. In *Proceedings of the 1st Cooking with Computer Workshop*, pages 29–34. Citeseer.
- James Pustejovsky. 2018. From actions to events: Communicating through language and gesture. *Interaction Studies*, 19(1-2):289–317.
- James Pustejovsky, Harry Bunt, and Annie Zaenen. 2017. Designing annotation schemes: From theory to model. *Handbook of Linguistic Annotation*, pages 21–72.
- James Pustejovsky, Eben Holderness, Jingxuan Tu, Parker Glenn, Kyeongmin Rim, Kelley Lynch, and Richard Brutti. 2021. Designing multimodal datasets for nlp challenges. *arXiv preprint arXiv:2105.05999*.
- James Pustejovsky and Nikhil Krishnaswamy. 2022. Multimodal semantics for affordances and actions. In *International Conference on Human-Computer Interaction*, pages 137–160. Springer.
- Michaela Regneri, Marcus Rohrbach, Dominikus Wetzels, Stefan Thater, Bernt Schiele, and Manfred Pinkal. 2013. Grounding action descriptions in videos. *Transactions of the Association for Computational Linguistics*, 1:25–36.
- Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. 2012a. A database for fine grained activity detection of cooking activities. In *2012 IEEE conference on computer vision and pattern recognition*, pages 1194–1201. IEEE.
- Marcus Rohrbach, Michaela Regneri, Mykhaylo Andriluka, Sikandar Amin, Manfred Pinkal, and Bernt Schiele. 2012b. Script data for attribute-based recognition of composite activities. In *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part I 12*, pages 144–157. Springer.
- Markus Rokicki, Christoph Trattner, and Eelco Herder. 2018. The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the international AAAI conference on web and social media*, volume 12.
- Victoria L Rubin and Elizabeth D Liddy. 2006. Assessing credibility of weblogs. In *AAAI spring symposium: computational approaches to analyzing weblogs*, pages 187–190.
- Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3020–3028.
- Hanna Schösler, Joop De Boer, and Jan J Boersema. 2012. Can we cut out the meat of the dish? constructing consumer-oriented pathways towards meat substitution. *Appetite*, 58(1):39–47.
- Alain D Starke, Martijn C Willemsen, and Christoph Trattner. 2021. Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence*, 4:621743.
- Christoph Trattner, Dominik Moesslang, and David Elsweiler. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1):1–39.
- Jingxuan Tu, Kyeongmin Rim, and James Pustejovsky. 2022. Competence-based question generation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1521–1533.

- Nicole Ummelen. 1997. *Procedural and declarative information in software manuals: Effects on information use, task performance and knowledge*, volume 7. Rodopi.
- Ielka Van der Sluis, Anne Nienke Eppinga, and Gisela Redeker. 2017. Text-picture relations in multimodal instructions. In *Proceedings of the IWCS workshop on Foundations of Situated and Multimodal Communication*.
- Ielka Van der Sluis, Lennart Kloppenburg, and Gisela Redeker. 2016a. PAT Workbench: Annotation and evaluation of text and pictures in multimodal instructions. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH) at COLING 2016*, pages 131–139.
- Ielka Van der Sluis, Shadira Leito, and Gisela Redeker. 2016b. Text-picture relations in cooking instructions. In *Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation: Proceedings of the Twelfth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 22–27.
- Ielka Van der Sluis and Hanna Mellema. Submitted. A recipe for success: The design, use and effectiveness of multimodal online baking instructions. *Multimodality & Society*.
- Ielka Van der Sluis and Gisela Redeker. 2019. The pat annotation model for multimodal instructions. In *6th European and 9th Nordic Symposium on Multimodal Communication*.
- Ielka Van der Sluis, Gisela Redeker, and Sannah Debreczeni. 2022. A text-based method to derive the main action structure in procedural instructions. In *AREA II: Workshop on the Annotation, Recognition and Evaluation of Actions held in conjunction with the 33rd European Summer School in Logic, Language and Information 8-19 August, 2022*.
- Ielka Van der Sluis, Renate Vergeer, and Gisela Redeker. 2018. Action categorisation in multimodal instructions. In *Proceedings of (AREA 2018)*, pages 22–27.
- Youri Van Pinxteren, Gijs Geleijnse, and Paul Kamsteeg. 2011. Deriving a recipe similarity measure for recommending healthful meals. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 105–114.
- Charlotte Vijfvinkel, Ielka Van der Sluis, and Gisela Redeker. 2018. I like to move it move it: Analysing first-aid instruction videos for moving a victim. In *TABU Dag 2018: The 39th International Linguistics Conference*.
- Janina Wildfeuer, Ielka Van der Sluis, Gisela Redeker, and Nina Van der Velden. 2022. No laughing matter!? analyzing the page layout of instruction comics. *Journal of Graphic Novels and Comics*, pages 1–22.
- Semih Yagcioglu, Aykut Erdem, Erkut Erdem, and Nazli Ikizler-Cinbis. 2018. Recipeqa: A challenge dataset for multimodal comprehension of cooking recipes. *arXiv preprint arXiv:1809.00812*.
- Asami Yajima and Ichiro Kobayashi. 2009. "easy" cooking recipe recommendation considering user's conditions. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, volume 3, pages 13–16. IEEE.
- Yu Zhang, Li Cheng, Jianxin Wu, Jianfei Cai, Minh N Do, and Jiangbo Lu. 2016. Action recognition in still images with minimum annotation efforts. *IEEE Transactions on Image Processing*, 25(11):5479–5490.
- Ziqi Zhang, Philip Webster, Victoria S Uren, Andrea Varga, and Fabio Ciravegna. 2012. Automatically extracting procedural knowledge from instructional texts using natural language processing. In *LREC*, volume 2012, pages 520–527. Citeseer.