Gastrolinguistics: Exploring Linguistic Uniqueness of Culinary Language in Modern Bengali Cuisine Corpus

Niladri Sekhar Dash
(Indian Statistical Institute, Kolkata)
Amrita Bhattacharyya
(Amity University, Kolkata)

The purpose of this paper is to record some unique and interesting linguistic features found in the Modern Bengali Cuisine Corpus (MBCC) and investigate how these features are represented in the cuisine texts of the corpus. The corpus is designed with curated cuisine texts obtained from various digital sources covering a small time window (2001-2024) with limited geographical variations. It contains only texts to reflect on some lexicosyntactic uniqueness of the text type. Since this is primarily a text corpus, the non-textual elements (e.g., *images, graphs, sketches, videos and emojis*) are removed from it to maximize textual representation and minimize linguistic inaccuracies. Subsequently, the corpus is put to some qualitative analyses to unveil some unique linguistic phenomena that are characteristically inherent to the culinary discourse. The initial analysis of the MBCC reveals the presence of cryptic syntactic structures, the prevalent use of imperative constructions, the high use of content words, frequent use of code-mixing and the strategies for targeted audience orientation. All these features, when considered together, underscore some inherent variabilities in linguistic structures dictated by communicative needs and manifest the remarkable proficiency of the human brain in extracting pertinent information from structurally complex texts. Further studies are necessary to understand the frequency of use of content words, distribution patterns of code-switching in discourse-diverse texts, linguistic strategies used in information encoding and the possibility of using this corpus in various domains of Bengali linguistics and technology developments.

Keywords: Gastrolinguistics, Bengali, Culinary Discourse, Cuisine Corpus, Linguistic Variability

1. Introduction

This paper opens an interesting line of inquiry through analysis of a corpus of cuisine texts (namely, the Modern Bengali Cuisine Corpus: MBCC) as an area of study of a natural language. It carries out some preliminary studies on the MBCC to understand the novelty and uniqueness of the Bengali language manifested in it. With evidence collected from MBCC, it shows that the language used in cuisine texts differs markedly from conventional Bengali texts, thus it prompts further inquiries into its distinctive linguistic manifestation, unique text composition patterns, peculiarities in text structure and mechanisms used to represent information (Barlow, 2011). The texts of the MBCC are full of content words, short and cryptic sentences, code-switching and imperative syntactic forms. The text becomes an object of investigation as it reflects on certain linguistic features and aspects that are hardly observed in regular modern Bengali informative and imaginative texts. The novelty of the text and the process of information embedding in the text make the MBCC an area of linguistic investigation to understand how people comprehend it and extract appropriate information without the support of discourse-level information.

The Modern Bengali Cuisine Corpus (MBCC) is compiled with cuisine texts produced in West Bengal, India and Bangladesh. The corpus is formed with sample texts of Indian Bengali recipes and sample texts of Bangladeshi Bengali recipes. At the time of collecting cuisine texts, it is observed that the Bengali language, which is used to compose cuisine texts, deviates from many standard linguistic norms in sentence formation, use of words in texts, code-switching and others. Thus, MBCC presents a unique set of features based on which the MBCC text can be distinguished from the texts of other genres (e.g., *literature, science, technology, mass media, judiciary*). The corpus also refers to some linguistic differences in the texts composed in West Bengal, India and the texts produced in Bangladesh. This area can be explored further in detail with larger corpora collected from the two different places. The primary purposes of this study are to identify the unique linguistic features noted in the MBCC and to investigate how these different linguistic features may be put against standard Bengali for further investigation.

It is often argued that the language of cuisine texts, with its unique jargon and text structures, is different from the language of a general text (Zhang, 2009). It provides new data and information to examine how unique linguistic strategies are deployed to compose cuisine texts and how people apply extralinguistic cues to comprehend these texts (Zhuoma, 2023). In general, the language of MBCC includes special vocabulary, special terms, and a selective set of verbs to describe the ingredients used in cooking, methods of preparation, types of flavours to be generated, courses to be followed to serve foods, processes to be followed to eat, and the kind of satisfaction the food eaters can have from a food item. In most cases, it includes a set of domain-specific technical terms, a set of descriptive words, a set of phrases, some code-switched expressions and a few borrowed words, which are either indigenous to the Bengali language and culture or borrowed from English and other languages and cultures (Ahmad *et al.,* 2023). Thus, linguistically, the MBCC exhibits rich and specialized linguistic features, which are largely different from the features noted in standard modern Bengali informative and imaginative texts. The present paper aims to address these aspects with reference to some data and information elicited from the corpus. The theoretical and applied relevance of this study are also addressed in different sections of this paper.

2. Major Research Questions

To meet the standards of an empirical corpus-based study of a language, it is necessary to define the research questions, provide a transparent and reproducible methodology, and present some quantitative findings with qualitative analyses. Since it is designed as a resource paper, it aims to present some information regarding the corpus documentation as well as some results derived from some concise pilot analyses. Elaborate results retrieved by statistical studies on the corpus may be used later to develop a 'focused study', where smaller sets of rigorously tested results will be presented to explore more of Bengali corpus linguistics to investigate how culinary discourse can be substantially useful for understanding the diverse linguistic dimensions of the language.

This paper aims to investigate the complexities noted in text formation and record the unique linguistic features of Bengali cuisine texts. Although these are not directly linked with explicit, testable research questions or hypotheses, this study brings in the Bengali cuisine corpus into the larger ambit of corpus linguistics as well as finds some potential linguistic differences of Bengali across the border of India and Bangladesh, which can be taken up as a potential and relevant area of study for the language with texts of other disciplines. Keeping this observation in the background, this study aims to address the following research questions with evidence collected from the MBCC:

(a) How does the language of MBCC differ from the language of the modern Bengali general text corpus?

(b) Do the sub-corpora of West Bengal and Bangladesh differ in the relative frequency of second-person honorific imperatives (e.g., *nin/din*) vs. inclusive futures (e.g., *niba/diba*)?

(c) Are imperative forms overrepresented in MBCC in relation to their use in the modern Bengali general reference corpus?

(d) How can code-switching and lexical borrowing be differentiated in the MBCC? How are these distributed across cuisine texts?

(e) How frequent are 'verbless' sentences in MBCC? Do they cluster in the list of ingredients used in cooking or appear in texts describing cooking procedures?

(f) How can the data and information derived from the MBCC be utilized in the sustainable development of community culture and heritage?

Since the MBCC is developed with some specific purposes, it is necessary to know the goals and motivations behind the creation of this corpus.

- Make a comparative study between the MBCC and the modern Bengali general corpus to find out how these two corpora of the same language differ on various linguistic features and properties (Ashari and Nugraha, 2022).
- Develop a separate lexical database relating to gastrolinguistics, which can be used to understand the discipline as well as translate food items and recipes across languages and cultures (Ahmad *et al.,* 2023).
- Start a corpus-based analysis on food-related borrowing words from English to Bengali and record their influences on reforming Bengali culture and language (Zhuoma, 2023).
- Generate a corpus-based sentence list relating to culinary, cuisine, recipe and gastronomy for machine learning and translation (Rungrueang *et al.,* 2022).
- Investigate how particular culinary terms of English have been translated into Bengali to build a regular lexical set to be naturalized as part of the regular vocabulary of Bengali speakers (Riyantika *et al.,* 2020).
- Develop a database of modern Bengali cuisine as an online platform with a collection of recipes with organized search and filtering operations (Lim, 2018).
- Help start-ups to develop food blogs and websites with a focus on indigenous Bengali cuisines, food items and recipes for making commercial products and services.
- Make the corpus available to researchers to study various aspects of food cultures, culinary trends and the linguistics involved therein.
- Explore how the modern Bengali lexicon is enriched with the addition of a large number of words and terms found in cuisine texts (Nordin *et al.,* 2013).
- Study how this new lexical database creates challenges for Bengali speakers with regard to their meanings, senses and usages.
- Investigate how different cuisines and culinary traditions have their own vocabulary and terminology that percolate into the Bengali language and culture.
- Analyze how cuisine-related terms reflect the unique ingredients and cooking methods of their respective regions and communities.
- Study how the modern Bengali life accommodates new culinary practices and thus opens up scopes for cultural expansion and inclusivity across cultures.
- Compare the vocabulary of cuisine texts with the values and priorities of the Bengali society. It is not just the vocabulary that matters; it is also the attitudes that are assimilated in the language of food and cooking.

As stated at the beginning, the primary aim of this study is to analyze the distinctive linguistic features of Bengali cuisine texts and to understand how these texts differ from the standard linguistic

practices of the language. By exploring the corpus, this study also desires to uncover the unique writing styles of content creators and understand the methods used to comprehend the texts by Bengali readers. The findings may contribute to a much broader understanding of how specialized texts are processed and interpreted by common language users.

3. **Theoretical Framework**

The concept of *gastrolinguistics*, which is proposed in the title of this paper, is a new idea that constitutes the features of both linguistics and gastronomy. It proposes to analyze the language of cuisine texts based on a corpus developed with data collected from varied cuisine texts. It desires to include all relevant methods and processes used in corpus linguistics, gastronomy and culinary studies. The coinage, at present, functions more as a label than an analytic framework, which the authors desire to develop in due course of time with the accumulation of new information, data and insights from analysis of various cuisine texts. It also desires to benefit from anchoring in established theories of register, genre variation and instructional discourse (e.g., *register analysis, information structure in procedures, politeness and directive strategies, content representation, machine learning of texts, text categorization anchors*) to constitute a workable model of gastrolinguistics for analyzing texts from gastronomy.

On the helm, the term *gastrolinguistics* is used in a lexicographic sense to refer to the study of the relationship between language, food and culture. The scope is wide enough to include the art and craft of preparing and serving food in societies, the esthetics of delicate and appetizing foods, cooking styles of particular geographical regions, the science of good eating, health-related benefits or hazards relating to particular food, and the ways people offer foods to their gods and goddesses. Against this background, this new portmanteau term desires to include the entire ecology of language, history, heritage, religion, philosophy, mythology, rites, culture, rituals, practices, ethics, aesthetics, culinary, cuisine, recipes, gastronomy and other aspects relating to cooking, food, community, knowledge, beliefs, heritage and other aspects of societies that are directly or indirectly linked with language, food and culture.

The term *recipe*, which is often used in language in a metaphoric sense rather than in a literal sense, refers to a set of instructions (or a series of instructions) for preparing a particular dish, including the list of ingredients required to prepare the dish. Etymologically, it is a medical term that was mainly used in medical prescriptions to refer to a formula or process for composing a remedial potion prescribed by a physician (1580s). In this paper, the term is used in its standard sense as recorded in gastronomy.

The term *cuisine* is broadly individualistic in sense as it refers to unique styles and methods of cooking of particular food items, keeping in view the characteristic features, geoclimatic properties of particular establishments, families, societies, regions, communities, or countries. Therefore, a *cuisine corpus* is theoretically meant to represent the texts that are produced with foods, cooking, rituals and other aspects specific to communities or societies as a central theme or content. The term is used in this paper in the sense defined here.

The *cuisine corpus* is not altogether a new idea in corpus linguistics; however, it is not a very frequent area of study of a language. In this paper, a *cuisine corpus* conceptually refers to a collection of written texts relating to cuisine. It includes texts relating to recipes, ingredients, cooking techniques, cultural contexts and more. It is a collection of special types of text of a language in digital form, which can be used for description, research, analysis and many other purposes of a language. The concept of *cuisine corpus* is different from the concept of *culinary text* with regard to inclusivity, composition and application. A *culinary text* usually refers to a text that teaches the process of cooking. As a part of an academic discourse where the science of food making is taught as a course to the learners, a culinary

text describes methods of cooking various foods and preparing various drinks based on specific contexts and situations. In the present scheme of study, the compound *'cuisine corpus'* refers to a structured collection of written text samples related to cuisine with proper reference to the issues and aspects like recipes, ingredients, cooking process, cultural significance, and others. Theoretically, a cuisine corpus can have the following applications for systemic analysis of language and societies:

(a) It is a useful resource for studying cuisine texts, understanding food cultures, analysing recipes and developing culinary applications.

(b) It supplies linguistic data and evidence to study words and terms used, the structure of syntactic constructions, the types of language deployed, the processes of information embedding and the nature of comprehension of texts.

(c) It situates different linguistic variations (e.g., *imperative vs. inclusive statements*) within larger models of sociopragmatic theories (e.g., *directive mitigation, honorificity, audience design*) to explain how variations are attested across text varieties.

4. Methodology Applied

To record the linguistic peculiarities of modern Bengali cuisine texts, the Modern Bengali Cuisine Corpus (MBCC) is compiled within a year (2025) with a short time window for the corpus data (2001-2024), keeping various logistic constraints in mind. It is made through a collaborative man-machine interface where scholars are engaged to collect, normalize and process corpus data mostly manually with the help of computers, Unicode-supported Bengali text and online Bengali texts processing tools. The corpus contains carefully curated cuisine texts sourced primarily from digital platforms. The short time window span ensures contemporariness of texts as well as confirms the relevance of the analysis of texts of a specific geoclimatic region of the Bengali language across a political border. The geographical diversity is minimized to focus on the texts of similar linguistic and cultural backgrounds. Each text is crosschecked by the research team to eliminate the presence of data or information (e.g., *noise, emoji, image,* etc.) not relevant to the present study.

4.1. Data Collection and Preparation

Since the primary focus of the research is on textual data, non-textual elements (e.g., *images, graphs, pictures, audio and video*) are not considered for the corpus. The team follows a uniform strategy by which the textual representation is maximized, while potential loss of information (due to the removal of non-textual elements) is minimized. Primarily, only those texts that are not supported with images, pictures and audio-visual information are taken into the corpus. Moreover, after initial compilation, the corpus is put under a series of text processing, which begins with creating metadata for the texts retrieved from various sources. The metadata, which has been developed for the MBCC, primarily relates to the cuisine texts with a focus on the name of dishes, recipes, ingredients, food type, food category, region and other information that help future users and system developers to understand and retrieve relevant text and information from the corpus. The metadata of the MBCC include the following key fields in alphabetical order (Table 1):

Table 1. Metadata Specially Developed for the MBCC

Key Fields	Metadata Information
Category:	snacks, breakfast, lunch, dinner, dessert, sweets, drinks
Cuisine type:	Bengali, English, Chinese, North Indian, South Indian, Continental
Dietary type:	Vegan, gluten-free, low-fat, veg, non-veg, Jain
Difficulty level:	Easy/ medium/ hard

Domain:	Cuisine/culinary
Duration:	Preparatory time/ cooking time/ total time.
Equipment:	Specific tools needed (e.g., wok, blender, oven, spoon, grill).
Ingredients:	List of items and quantities
Instructions:	Step-by-step preparation method
Name of creator:	Creator of the recipe
Name of dish:	Name of the dish
Narration:	1st person/ 2nd person (imperative)/inclusive
Nutrition:	Calorie count/ fat content, etc.
Region:	West Bengal/Bangladesh
Religion:	Hindu/ Muslim/ Christian/ Sikh/ others
Consumer:	One/two/many
Description:	Elaborate/brief/cryptic/ metaphorical
Text type:	Text/ text & image/text & audio/ text, audio & video
Time span:	2001-2024
Usage type:	Formal/ informal/ personal/ familial/auspicious

Since image, audio and video data are not included in the MBCC, the metadata relating to image, audio and media texts are not taken into consideration in the list. However, this remains an area of serious study for those who will be interested in collecting image, audio and video data of cuisine texts, where special metadata information will be required for audio texts, images, image descriptions, and other aspects of multimodal texts. The entire text data of the MBCC is collected from some free and open platforms of the cyberspace to ensure that no legal and copyright issues are associated with its acquisition, analysis, processing and utilization. The collected data are preserved in Rich Text Format (RTF) and .doc format with UTF-8 encoding. The text data is processed with the help of Notepad++, segmented at the segment and sentence level, and populated in Excel files for the addition of metadata and analytical information.

4.2. Regional Distribution

The MBCC is made of two sub-corpora, so to speak. One subcorpus contains texts of cuisines that are composed by the creators of West Bengal, India. This subcorpus contains data from 5 different sources; it includes 530 recipes, 17294 sentences and 135819 words (Table 2). For the purpose of reference, this subcorpus is named as West Bengal Cuisine Corpus (WBCC).

Table 2. Composition of West Bengal Cuisine Corpus (WBCC) of MBCC

Name	Recipes	Sentence	Word
Boldsky	126	4,037	31,880
Cookpad	43	1,518	9,353
Ei Samay	119	8,442	58,981
Hyangla Heshel	191	2,210	26,688
Pratidin	51	1,087	8,917
Total	530	17,294	135,819

The other subcorpus contains texts of cuisines that are composed by creators from Bangladesh. This sub-corpus contains data from 6 different sources; it includes 239 recipes, 6994 sentences and 42785 words (Table 3). For the clarity of reference, this sub-corpus is named as Bangladesh Cuisine Corpus (BCC).

Table 3. Composition of Bangladesh Cuisine Corpus (BCC) of MBCC

Name	Recipes	Sentence	Word
Cookpad	87	2,004	12,875
Bangladeshi Kitchen	72	2,237	13,628
Banijya Pratidin	33	1,078	7,132

BD News 24	23	719	3,579
Shajgoj	14	408	2,511
Eisamay	10	548	3,060
Total	239	6,994	42,785

Besides these two major sub-corpora, some more cuisine texts are also collected from some digital sources whose identities are quite obscured. There are nearly 111 recipes, 2723 sentences and 16454 words. For reference purposes, this corpus is called 'miscellaneous', and its data is taken into analysis. Taken together, the MBCC contains, in total, 880 (< 530 + 239 + 111) recipes, 27011 (< 17294 + 6994 + 2723) sentences and 195058 (< 135819 + 42785 + 16454) words.

It has not been possible to follow any specific sampling criteria for compiling the MBCC due to many logistic constraints, some of which are as follows:

- The MBCC is dominated by data generated by some specific food blogs and websites (e.g., *Hyangla Henshel* from West Bengal and *Cookpad* from Bangladesh) because most of the data are available from these sources.
- The BCC is smaller in size as the recipes are mostly found in video format, which grossly outnumber the recipes found in plain text format.
- There have been restrictions on access to certain websites due to different geographical locations and copyright issues.
- Many Bengali recipes are written in English, rather than in Bengali. These are not taken as they are not relevant for this corpus.
- Some recipes have only audio and video representation without any text. These are also not taken into account.
- Many recipes include images of different steps of preparation and methods to describe the cooking procedure. They have a very limited amount of text to be considered.

Due to the very limited source of data, no inclusion or exclusion criteria for texts are adopted for this corpus. Moreover, no linguistic rules or conditions are applied in the selection of texts. Anything related to the Bengali cuisine and written in the Bengali script is taken into account. Maintaining balance with regard to sites, subgenres (*e.g., home blogs, food news, tourist blogs, food portals, food-related advertisements*) has not been possible due to the very limited scope for data availability. However, marking the region of origin of text (*text-WB vs. text-BD*) is noted in the metadata to avoid site-based style variations.

In essence, whatever text samples are available are collected. Since the removal of images from texts, in principle, can affect and alter the basic character and load of information of a text (as visuals and layout often encode essential procedural information), primarily those texts, which are not supported with images, audio, or videos, are collected. Therefore, for the present study, it is not required to record, for each recipe, the details of the removed elements (e.g., *images, photos, diagrams, tables*) as well as tag each section and subsection of visual elements (e.g., *title, ingredients, method, notes, tools,* etc.). Different structural markup mechanisms (e.g., *TEI encoding*) may be considered when the visual elements will be included in the future multimodal cuisine corpus.

5. Data Analysis: Linguistic Aspects of the MBCC

The data collected in the MBCC is put to both quantitative and qualitative analyses. Initial findings record the high use of cryptic sentence structures, use of imperative constructions, high use of content words and different patterns of code-switching. During analysis, attention is paid to the factor of orientation of a text designed for a targeted audience and the processes that are used to represent

information in a text. These features are examined to determine their roles in the application of specific linguistic structures in cuisine texts and to measure the degree of communicative efficacy of cuisine texts on the target readers. Some initial findings that exhibit some unique linguistic characteristics are discussed in the following subsections.

5.1. Use of Verbless Sentence

The use of a verbless sentence is an attested feature of modern Bengali text. It implies that in Bengali, a grammatically well-formed and semantically valid sentence can be formed without the use of a finite verb. Although this kind of construction is quite frequent in Bengali speech, it is not so recurrent in written Bengali texts. Earlier scholars argued that verbless sentences in Bengali are possible only in the simple present tense, where the absence of a finite verb or a copula is automatically taken care of by the tense itself (Chaki, 1996; Thompson, 2010). In reality, however, there are instances in the modern Bengali texts corpus where verbless sentences are found to be used in the past and perfect tense (Dash, 2017). The present MBCC shows a different kind of verbless construction, where the lists of ingredients, which are supposed to be used in a recipe, are arranged in a linear order to produce a construction that looks like a sentence but is not a sentence in the true sense of the term. It is a unique example, where several clauses are arranged in a sequential order with subtle use of commas to form a large syntactic form, where each clause is constructed without a verb, although each one is eligible to be reframed as a full sentence with one or more verbs. This kind of construction is hardly noted in general Bengali texts, as far as the data of a modern Bengali corpus shows (Dash, 2005). The following example (1a) shows how a large pseudo-sentence is formed in Bengali cuisine text by joining several clauses (for better comprehension, the example is provided in original Bengali script followed by *Indic Roman with diacritics* and a running translation in English), to serve the purpose of the text creators.

(1a) ekṭā kaci bāṅśer khol, ādh keji murgir māṃsa, pẽyāj kuci, ādā bāṭā, rasun bāṭā, ṭ ameṭo kuci, halud gũṛo, jire gũṛo, kayekṭā kācā laṅkā, sarṣe tel, svād anuýāyi nun o cini, ār ekṭā phayel.

(1b) "One tender bamboo bark, half a kilogram chicken, chopped onion, ginger paste, garlic paste, chopped tomato, turmeric dust, cumin dust, a few green chilies, mustard oil, salt and sugar to taste, and one foil".

The non-use of a verb in this kind of pseudo-syntactic construction has a purpose. It implies that since each of the clauses requires the same type of verb to be used to form full sentences, and since repetition of the same verb form may affect the precision of information represented through the clauses vis-à-vis the sentence, the text composers have willfully eliminated the verb from the sentence. Surprisingly, it does not affect much to the information the text composers want to share with the target readers. This is very much possible because the primary goal of cuisine text composers is to share information, guide target readers and instruct prospective audience to adopt and apply the knowledge for their own purposes.

5.2. Cryptic Syntactic Structure

The use of short and cryptic syntactic constructions is quite predominantly observed in the MBCC. The phenomenon is quite interesting and captivating. In a general sense, the use of this kind of sentence has a purpose. The simple argument is that most of the sentences in the MBCC are imperative in sense and imperative sentences, due to their unique purposes, are often short in length and cryptic in form. Because of this reason, the sentences in MBCC are mostly short and cryptic. With regard to

the number of words in a sentence, imperative sentences are formed within 5 to 10 words (2a-6b), which are shorter in length than other types of sentences (e.g., *descriptive, indicative, interrogative*). Perhaps shorter sentences generate precise information and a stronger impact on the target audience. In essence, specific motivational factors may be instrumental behind the formation of short and cryptic sentences in cuisine texts, as the following examples show (2a-6b).

(2a) duṭi mājhāri ālu ardhek kare keṭe nin.
(2b) "Cut two medium-sized potatoes into halves."
(3a) garam garam paribeśan karun.
(3b) "Serve hot."
(4a) beśi kare sarṣe tel mākhiye din.
(4b) "Smear some extra mustard oil."
(5a) jvalanta ununer upare rākhun.
(5b) "Put on a burning oven."
(6a) kayekṭā ālu greṭ kare nin.
(6b) "Grete a few potatoes."

The above examples (2a-6b) show that all the sentences are made with fewer than ten words, a feature which is hardly noted in other kinds of sentences in Bengali (Dash, 2005). This can be treated as a marked feature of Bengali cuisine texts.

Another interesting feature is noted when texts of the two Bengali communities across the border are compared. Compared to the texts produced by the Bengali bloggers living in West Bengal, India, the length of texts produced by the bloggers living in Bangladesh is longer in terms of the number of sentences and words. For instance, for a particular recipe (e.g., *preparing a mixed vegetable curry*), the length of a piece of text produced by a West Bengal blogger is much less in number of sentences and words (say, 20+ sentences and 200+ words) than the text produced by a blogger from Bangladesh (say, 30+ sentences, 400+ words). The linguistic factors behind this difference are not yet known. Perhaps one has to go beyond the text to understand the reasons behind this difference, with an analysis of texts from other disciplines.

5.3. Imperative Constructions

The predominant use of direct instructions is another marked feature of the MBCC texts. This feature is more predominant in the texts produced in West Bengal (7a -11b), where repeated use of the verb form *nin* "you (+hon.) take" (a 2nd person (Sg.+Pl.) honorific imperative verb form) refers to the invisible presence of distance between the writers and the readers. Anyone who reads the text understands that the text creators knowingly try to keep a distance from the target readers, thereby maintaining a relational distance clearly noticeable from the use of honorific verb forms. This is a typical and regular strategy of language use in Bengali, where speakers maintain a distance from hearers by deliberate use of honorific verb forms. Here, the text creators are instructing the target readers to do the work, whereas they themselves are not involved in the process, as the following examples (7a-11b) show:

(7a) gā-mākhā hale nāmiye nin.
(7b) "Keep it down when it is smeared on the body."
(8a) nārkelṭā prathame jale beś kichukṣaṇ ḍubiye nin.
(8b) "Keep the coconut in water for some time."
(9a) erpar nārkeler sādā aŋśaṭā chāṛiye nin.
(9b) "After that, peel off the white part of the coconut."

(10a) sarṣe candaner mato mihi kare beṭe nin.
(10b) "Make a paste of mustard seeds like a sandalwood paste."
(11a) pẽyāj ekebāre jhirijhiri kare keṭe nin.
(11b) "Cut onions into thin and tiny pieces."

On the other hand, in the cuisine text produced in Bangladesh, one can easily notice the more frequent use of descriptive sentences. In most cases, these are 'inclusive constructions' as the verb forms (e.g., *rākhba* "I/we shall keep", *niba* "I/we shall take", *diba* "I/we shall give") include both the content writers and the target readers. These verb forms are 'inclusive' in the sense that they indicate an act where the speakers and others, who are referred to in the sentence, are included as agents of the act. In this case, the content writers, while giving instructions to the target readers about the recipe, are also including themselves as members of an imaginary group engaged in a particular process of cooking, as the following examples (12a-16b) show. A close look at the examples shows that these finite verbs are primarily inclusive in nature, where speakers are involved in the process of an action along with the target readers.

(12a) cāl bhālo kare dhuye pāni jhariye rākhba.
(12b) "(We) shall wash rice and clear its water."
(13a) hãṛite parýāpta parimāṇe pāni phuṭiye niba.
(13b) "(We) shall boil enough water in a pot."
(14a) tāte labaṇ o ek cāmac tel diba.
(14b) "(We) shall give salt and one spoonful of oil."
(15a) cāl diye neṛe niba.
(15b) "(We) shall stir with rice."
(16a) hālkā kare bheje niba.
(16b) "(We) shall fry it light."

Further investigation is needed to understand why the two types of constructions are used in two types of Bengali found across the border. One reason might be as follows: the content creators of Bangladeshi cuisine texts are predominantly immigrants residing abroad. They adopt, for various sociocultural reasons, an inclusive approach while writing cuisine texts, and the use of inclusive verb constructions (e.g., *āmrā diba* "we shall give", *āmrā karba* "we shall do", *āmrā niba* "we shall take") creates a kind of ambiance where the content writers feel themselves included with their intended readers. On the other hand, the content writers from West Bengal, India, are primarily residents of West Bengal. Instead of adopting an inclusive strategy, they follow a strategy of using formal and honorific verb constructions (e.g., *āpni karun* "you do", *āpni nin* "you take", *āpni din* "you give"), which they think are more impactful in capturing the attention of their target readers. The use of 2nd person neutral verb form (e.g., *tumi dāo* "you give") is hardly found in the texts produced in West Bengal. Perhaps, a far deeper analysis of the socio-pragmatics and discourse is required to shed light on the linguistic and extralinguistic factors responsible for showing such differences in two varieties of Bengali.

5.4. Density of Content Words

The MBCC texts are featured with a density of domain-specific content words. The analysis of the corpus shows that a large number of content words are used in the texts, the majority of which belong to the jargon of the discipline (Table 4).

Table 4. Sample Text with a High Density of Content Words in MBCC

Bengali Cuisine Text: Text in Indic Roman with Diacritics
Total Number of Words: 66; Content Words: 50 (75.75%)
nārkelṭā[NN] prathame jale[NN] beś[ADJ] kichukṣaṇ[NN] ḍubiye[NFV] rākhun[FV]. ete korānor[NN] samay[NN] nārkeler[NN] khayeri[ADJ] aṅśa[NN] nārkel[NN] korār[NN] madhye mile[NFV] ýābe[FV] nā. erpar nārkeler[NN] sādā[ADJ] aṅśaṭā[NN] chāṛiye[NFV] beṭe[NFV] nin[FV]. miksite[NN] mihi[ADJ] kare beṭe[NFV] nite[NFV] pāren[FV]. ābār cāile[NFV] kuriyeo[NN] nite[NFV] pāren[FV]. bāṭār[NN] samay[NN] kintu ete jal[NN] deben[NN] nā. Ebār kaṛāi[NN] garam[ADJ] kare, khub[ADJ] kam[ADJ] āce[NN] cini[NN] ebaṃ jal[NN] diye[NFV] sirā[NN] tairi[NN] karte[NFV] śuru karun[FV]. ete miśiye[NFV] nin[FV] thẽto[ADJ] karā elāc[NN].

The table contains a very small sample of a text containing just 66 words, out of which 50 words (75.75%) are content words (tagged with IDs of various parts-of-speech). A similar kind of count on a general Bengali text shows a lower percentage of presence of content words (Mallick *et al.,* 1998). There are, however, limited varieties among the content words used in this kind of text. Majority of them include the names of ingredients (e.g., *nārkel* "coconut", *cini* "sugar", *elāc* "cardamom") and instruments (e.g., *ãc* "oven", *miksi* "mixer", *kaṛāi* "wok") used for preparing some dishes, a set of verbs defining the nature of actions carried out (e.g., *ḍubiye rākhun* "sink in", *miśiye nin* "mix up", *śuru karun* "start"), a set of adjectives specifying the quality of items or processes (e.g., *khayeri* "purple", *garam* "hot", *kam* "light") and a few proper names relating to the materials and dishes.

5.5. Code Mixing

The frequent integration of multiple languages or dialects is a regular and prominent feature of this kind of text. This is also a clear indicator to understand how Bengali life and society are open to assimilating external cultural traits and practices. The assimilation of foreign food items and practices is a clear indication to realize that the traditional Bengali foods and practices are gradually losing their ground. In the MBCC, the presence and predominance of English words and expressions are examples of code-mixing or borrowing, which are used as useful strategies for describing various cooking states/procedures, as the following examples (17a-28b) show.

(17a) cikenṭā bhālo kare dhuye nin.
(17b) "Clean the chicken well."
(18a) ekṭā pātre cikener pisguli rākhun.
(18b) "Keep the pieces of chicken in a plate."
(19a) kyāoṛār essens o garam maślā meśān.
(19b) "Give the essence of Kewra (Pandanus) and cardamom."
(20a) jāfrān miśiye myārineṭ kare nin.
(20b) "Marinette it after mixing with saffron."
(21a) ebār frije du ghaṇṭār janya rekhe din.
(21b) "Now keep it in the refrigerator for two hours."
(22a) rānnāṭā myārineṭer upar nirbhar karche.
(22b) "The cooking depends on marination."
(23a) ekdam kam sime rānnāṭā karte habe.
(23b) "(You) have to cook in a very low sim."
(24a) ebār cikener pisgulo diye din.
(24b) "Now give the chicken pieces."
(25a) myārineśaner samay seguli diye din.
(25b) "Give them during the time of marination".
(26a) jānben reḍi haye geche cāp.
(26b) "Be sure, the chap is ready"

(27a) paribeśan karun resṭurenṭ sṭāil ciken cāp.
(27b) "Serve the restaurant-style chicken chap."
(28a) jame ýābe pārṭi.
(28b) "The party will be a great hit."

There are debates whether the use of English terms like *chicken, frying, marinate, mixing, grilling, serve, party, sim, essence, restaurant* and others in modern Bengali cuisine texts should be treated as code mixing or borrowing. They may be treated as borrowings as they are quite frequently used in Bengali cuisine texts through the regular process of lexical naturalization in Bengali (by way of using Bengali affixation and inflection with these words). On the other hand, they may be treated as examples of code-switching, as many of the terms are not yet accepted as regular Bengali words to be included in standard Bengali dictionaries. Until and unless these are included in standard Bengali reference dictionaries, it is better to consider them as instances of code-mixing. The final decision, however, depends on the inferences drawn from analysis of large, multi-disciplinary and representative Bengali text corpora, which can be an interesting area of investigation for future Bengali language researchers.

5.6. Imperative Mood and Tense Variations

The composition of a cuisine text is formal with a structured format including title, ingredients and cooking procedure, which indirectly indicates a definite time of an action (e.g., *past, present, or future*) denoted by the use of finite verbs. A preliminary analysis of the MBCC shows the use of three types of finite verb forms, namely the following (Table 5):

(a) **Type 1:** Verbs indicating imperative mood. The content creators have used those verb forms, which indicate imperative mood. It is one of the most common patterns of formation of Bengali cuisine texts. The imperative mood, by default, always uses verbs in the simple present tense (Type 1).

(b) **Type 2:** Verbs indicating simple future tense. The content creators form every sentence of a text in the future tense, as if to inform that the mission is still unaccomplished and the readers have to take part in it with all possible preparations and resources (Type 2).

(c) **Type 3:** Verbs indicating simple past tense. The text creators intend to inform that the process of cooking is complete, the readers need to learn and assimilate how it is done by the text creators. A sense of completion of the event is expressed by the use of verb forms denoting past tense (Type 3).

Table 5. Imperative vs. Future vs. Past Procedural Narratives in the MBCC

Types of Finite verb Forms		
Type 1	Verbs indicating imperative mood.	bāniye phelun, jhariye nin, kare rākhun, mākhiye nin, ḍhele din, muṛiye nin, basiye din, rākhte pāren, kare din, dekhe nin, nāmiye nin, bheje nin, phuṭiye nin, diye din, nāmiye phelun, mākhiye rākhun, seddha karun, tule nin, bhājte thākun, miśiye din, ḍhākā din, khule dekhun, paribeśan karun
Type 2	Verbs indicating simple future tense.	keṭe nite habe, bhājte habe, nāṛte habe, tule dite habe, rākhte habe, phoṭāte habe, nāmāte habe, nāṛācāṛā karte habe, pheṭāte habe, ḍhālte habe, dhuye nite habe, bhijiye nite habe, chāṛiye nite habe, beṭe nite habe, greṭ kare nite habe, mekhe nite habe, maṇḍa kare nite habe, nāmiye rākhte habe, paribeśan karte habe
Type 3	Verbs indicating simple past tense.	bāniye phellām, beṛe phellām, bheje nilām, pesṭ diye dilām, rānnā karlām, kaṣiye nilām, kuṭe nilām, khosā chaṛālām, kuciye nilām, tule rākhlām, jhariye nilām, ḍhele nilām, phoṛan diye dilām, nāṛācāṛā kare nilām, jal diye dilām, bandha kare dilām, kamiye dilām, miśiye nilām, paribeśan karlām

The most striking aspect of this text is that there is hardly any instance of mixed use of all three types of verbs in a single recipe. That means, if a particular recipe is created with verbs indicating imperative mood, the entire text is made in this manner without using verbs of future or past tense. The same is equally true of the texts that have been created using the other two types of verbs. Another interesting observation may be made with regard to the use of the three different forms of verbs in cuisine texts. Pragmatically, the verb forms differ based on the kinds of rapport the content creators want to develop with the target readers. Based on their goals, discourse motives and expectation for acceptance of their texts by target readers, the content composers usually use different forms of verbs.

5.7. Other Characteristics

Besides the use of unique linguistic characteristics discussed in the previous subsections, the MBCC reveals some other characteristics, including the following. These features are, however, not elaborated here due to a lack of space.

- Heavy use of jargon relating to food and cuisine (e.g., braising, sautéing, grilling, risotto, recipe, temper, pantry, cup, defrost, frying pan, sausage, sim).
- Descriptive types of text (as every cuisine text describes how a food item has to be processed, prepared, cooked, and served).
- The use of onomatopoeic expressions. The content writers use these terms to express taste, texture, appearance and evoke sensory experiences.
- The use of unique ingredient and dish names (e.g., mirepoix, chiffonade, al dente, au gratin, bloody mary).
- The use of place names linked with specific food items, e.g., *nijāmī ciken* "chicken prepared at the Nizam", *jayanagarī moyā* "moya sweets made at Jayanagar", *ḍhākāi paroṭā* "a kind of bread made at Dhaka", *kāśmirī polāo* "a kind of rice made in Kashmir", *ciken āphgānī* "a kind of chicken prepared in Afghanistan", *hāydrābādī biriyani* "biriyani produced at Hyderabad", *mālābārī ciken* "kind of chicken prepared at Malabar", *cāinij cāṭni* "a tasty desert made in China", *pāṭnāi hāluyā* "a sweet dish made at Patna", *moglai paroṭā* "a kind of bread prepared during the Mughal era", etc.) and others.

6. Major Findings and Discussions

The MBCC not only contains recipes that are typical to the Bengali community but also contains many recipes, which are taken from a wide variety of societies, cultures, regions and countries. Besides recipes of different parts of India, it also contains recipes of many countries and communities of Europe, America, Australia, Africa and others. Moreover, it also contains recipes of many indigenous, ethnic and tribal communities as well as old and forgotten recipes as a part of ancestral heritage.

An intimate look into the MBCC shows that different cuisine texts and culinary descriptions possess some vocabulary and terminology, which are somehow different from other kinds of modern Bengali texts. They collectively reflect on the unique ingredients and cooking methods that are prevalent in the modern Bengali society across the border. In this context, some discussions regarding cultural significance and shaping the identity of the Bengali community at large may be summarized in the following manner.

- The MBCC is a useful resource for promoting diversity and inclusivity. It exhibits how Bengali people are open to accepting cultural practices of other societies and communities by promoting them for the Bengali community. It shows how the Bengali community assimilates different

cultures to broaden understandings of diversity. By exploring diverse cuisines, the content creators give insights into various societies' history, traditions and social structures.

- Cuisine texts are a good source of experiencing and learning about different food types that exist between and within specific communities and cultures. People cook the same food item or dish in different ways with different ingredients. From the varieties recorded in the MBCC, it is possible to learn many things about a particular individual or group's history and culture. On many occasions, the process of cooking the same dish is found to vary based on experience, process, ingredients, or climate. Many unknown stories of Bengali society, individuals, locality and culture are embedded within such cuisine texts.

- The MBCC silently promotes inclusion. It shows how a particular food item, which is not a part of the life of the Bengali community, is cooked, prepared and savored by the community. It is really fun when the Bengali community learns the method of cooking an item that has never been a regular dish for the community. For instance, in earlier days, learning to prepare *idly* and *dosas* was an event of great fun in Bengali families. At present, it is fun to prepare *pizza* or a *burger* at home. Learning a new dish is almost synonymous with honouring a different culture and practice. By using cuisine texts, it is possible to be culturally informed and to build positive relationships with other cultures and societies.

- The cuisine texts play an important role in shaping the Bengali culture. To learn to prepare a new dish from a different culture is an important practice of expanding the Bengali culture. It helps the Bengali community assimilate different cultural values, reproduce those values in their life over generations, and build an inclusive identity. On the other hand, the recipe for preparing a food in a different manner is an important part of an ethnic identity. For instance, different recipes for preparing biriyani at Srinagar, Delhi, Lucknow, Calcutta, Hyderabad, Mysore and Trivandrum in India are a mark of unique cultural identity. In most cases, people take extra efforts to protect the recipe from competitors to preserve their distinct cultural identity.

- The cuisine texts of the MBCC are characteristically different from political, commercial, or media texts. They reflect on those aspects of social life through which the cultural beliefs and practices of the Bengali community are distinctly expressed. For example, like other communities, the Bengali community also knows which foods are good and healthy, the time and process of preparing particular food, foods that are not permitted during specific events or occasions, foods that are harmful to consumers during specific seasons, foods that are not meant for children or elderly people, food that are to be prepared during specific religious or auspicious events, the foods that cannot be offered to guests, food and drinks that can be offered to gods and goddesses, and many such issues (Dash *et al.,* 2020). In fact, many of the cuisine texts indirectly guide the Bengali community to follow the 'food rules' or 'food norms', which are a kind of knowledge that is transferred through generations.

- Like other communities, the Bengali community also makes distinctions between 'low food' (*unsophisticated, not to be served at formal occasions*) and 'high food' (*sophisticated, to be served at formal events*), which are historically aligned with geography, climate, agriculture, ecology, culture, religion, custom, myths, belief, disease and other socioeconomic factors of the community. Also, it links with those Bengali people who have accumulated a lot of knowledge about the Bengali cuisine and have the authority to define which foods are worthy of respect and to be disparaged.

- The cuisine texts often serve as a medium for social and cultural exchange within and across the Bengali community. With their use, community people experience elements of cultures that are reinvented for them. Through these texts, the community becomes receptive to new and novel experiences and seeks opportunities to experience foods associated with different regions of the community. They value them as these experiences enrich their lives. When people of one

region prepare and eat food from another region, it becomes a new way of sneaking into others' lifestyle and culture, and it results in gathering new experiences outside their one culture. It can probably make them modest and tolerant in understanding the cultural differences of the same community across geographical regions.

- The language used to describe the preparation and presentation of food items reflects finer cultural attitudes and values of the Bengali communities. In the case of cuisine texts, language plays an important role in establishing the relationship of the community with the food they eat. For example, although food etiquettes are not included in cuisine texts, there are cues about the courses of action to be followed during the time of taking a food item or a series of food items. For the Bengali community, starting with a dish of bitter taste (e.g., *śukto* "a dish made with vegetables, primarily with bitter guard") and ending with a sweet dish (i.e., *pāyas* "a sweet dish made with rice, milk and sugar") is a common practice, while for the Tamil community, starting with a sweet dish and ending with a sour dish (e.g., *curd rice or rasam rice*) is an established pattern. This kind of information is not very uncommon in the MBCC, where the content writers mention when and how to start or end a particular course.

- The corpus also offers data and information about gender roles and power dynamics from the ways in which the cuisine texts are prepared and made available. Starting with the culinary traditions, the roles of male and female members are quite different in this case, too. It is noted that the majority of texts are prepared and produced by female writers, which indirectly indicates that the task of cooking is primarily a job of the female members, at least as far as the present cuisine corpus is concerned. It implies that preparation of food is a primary job of the female members in the Bengali community, and due to their longer exposure to this field, they are better individuals to write texts on this topic. By way of showing how food items are prepared and consumed, they share significant information about the role of gender in society and the power dynamics of the modern Bengali society. The feminism associated with this practice is a statement of the gender dynamics of the Bengali society, established and accepted for generations. It can be an interesting study to explore this area to know the ways in which food preparation has been an active agent in the history of Bengali culture to stereotype feminism with cooking.

A cursory look into the MBCC reveals that it contains much interesting information for corpus researchers. It carries information of 'recipes' (i.e., *detailed instructions on a sequential order for preparing dishes, including ingredients, quantities and cooking methods),* 'ingredients' (*lists of items and spices used in specific cuisines, their properties, regional variations and the processes of using them in a particular dish*), 'cooking techniques' (*information about methods and ways of cooking and their application in various cuisines*), and 'cultural content' (*information about the cultural significance of particular food and dishes within specific regions and communities*), which can be investigated further to carry forward the new slogan: 'corpus analysis for social purposes'. On this line of thought, the MBCC is a valuable resource for various purposes, including the following:

- **Research:** Analyzing culinary trends, identifying regional variations and understanding the evolution of food culture (Heidari and Jamalzadeh, 2020).
- **Development:** Creating recipe recommendation systems, developing new dishes and designing culinary experiences.
- **Education:** Providing resources for learning about different cuisine types and techniques, as well as their cultural significances.
- **Translation:** Translating recipes and culinary terms for the purpose of developing jargon databases for the language (Zhang, 2009).
- **Linguistics:** Understanding the changes taking place in linguistic forms and structures,

borrowing of foreign words and terms into language, nature and patterns of code mixing, assimilation and naturalization of domain-based concepts and system adaptation (Dash *et al.,* 2009).

7. Present Impact and Future Implications

The MBCC contains a domain-specific text, which is different in type, composition and character. It includes texts from food blogs, cooking instructions, food descriptions, recipe details and culinary narratives. Taken together, it represents an amalgamation of ideas, information, objects and practices that are unique and representative of a specific type of text, society and culture (Popovski *et al.,* 2019). It is important to understand the form and content of cuisine texts to critically examine the complex structure and representation of information relating to a speech community. It is a useful resource, the multidisciplinary analysis of which can help to comprehend a community as well as represent the community on global platforms. A critical analysis of MBCC can help to preserve cultural identity, foster social connectivity, pave ways for cross-community interchange, make roads for cultural assimilation and influence sustainable economic development.

The MBCC not only reports on traditional Bengali recipes, cooking rituals, food preparation, food processing and cooking methods, but it also shows how cooking foods and related activities can become integrated parts of the Bengali community in the process of preserving its unique cultural identity, cultural practices and cultural heritages. In the multicultural Bengali society, it can promote concepts like *community kitchen, food festival, collective cooking*, which can be instrumental in promoting inter-community co-existence, mutual cooperation and community cohesion through shared methods and collective meals. Moreover, a critical study of the MBCC may help to realize its impact on cultural identity and preservation of Bengali cuisine knowledge in the following ways:

- **Transmit cultural knowledge:** The MBCC, like cookbooks and recipes, is a repository of Bengali cultural knowledge handed to the present generation for practice and preservation and to be transferred to future generations.
- **Strengthen cultural identity:** The foods, the drinks, the recipes and the activities relating to these have been interwoven with the Bengali cultural identity over generations. The MBCC can help the Bengali people and societies connect with their heritage and maintain a sense of belonging.
- **Promote cultural exchange:** Through sharing culinary texts, the MBCC facilitates cultural exchange and understanding between the Bengali communities living across the border and fosters appreciation for diverse food traditions across geoclimatic areas.
- **Facilitate social gatherings:** Food is one of the indispensable elements in social events and celebrations. The MBCC offers scope for sharing culinary experiences to strengthen social bonds across families and communities.
- **Foster community building:** The MBCC encourages offline cooking classes, food festivals and community kitchens. It also encourages Bengali people to connect with each other, learn from each other and build stronger relationships across the border.
- **Promote social inclusion:** By highlighting different culinary practices, the MBCC promotes understanding, peer-habitation and inclusivity in the multicultural Bengali societies.
- **Support local food industries:** The MBCC promotes consumption of local foods to support local farmers, food producers and restaurants.
- **Promote food tourism:** By including unique cuisine texts, the MBCC likes to attract tourists interested in experiencing Bengali foods and boost the local economy (Laosrirattanachai and Ruangjaroon, 2020). Tourist agencies can use MBCC to refer to ethnic Bengali foods to help

economic growth by supporting local foods and tourism.

- **Create economic opportunities:** The MBCC may encourage food and beverage industries to create job opportunities for cooking Bengali foods to help the economy of the region (Laosrirattanachai and Ruangjaroon, 2021).

On the other hand, the MBCC can play a crucial role in shaping the Bengali societies by way of preserving cultural identities, fostering social connectivity, enriching relational bonding and leading to economic development through preservation, promotion and sharing of food cultures and practices. Also, it has the potential to trigger some innovative research activities, such as the following:

(a) **Frequency of content words:** analyzing usage patterns and the significance of content words in the MBCC can be an area of study to investigate why content words are heavily tilted in the cuisine corpus and what purposes they serve in the text.

(b) **Patterns of code-switching:** This area has not been much explored with data of real-life language use. The MBCC is a good resource to study how code-switching and code-mixing are used in Bengali cuisine texts, how these are distributed across different types of cuisine texts and how the basic rubrics of the matrix language framework (Myers-Scotton, 1983; Myers-Scotton and Jake, 2001) are manifested in Bengali cuisine texts.

(c) **Information encoding:** The MBCC is a good resource for examining the linguistic techniques and methods used to encode and convey information in cuisine texts.

(d) **Impact of visual elements:** The MBCC contains only textual data without any visual component. Therefore, it is a good resource for assessing how the exclusion of visual elements can affect the retention of information and comprehension of a text.

(d) **Technolinguistic applications:** The MBCC can be used to compile a cuisine-related lexical database annotated with multi-layered metadata. It will facilitate applications in machine learning (*algorithms for text analysis and generation*), multimodal text processing (*integrating text with other modalities such as speech and image*), automated text-to-speech systems (*improving accuracy and naturalness in speech synthesis*), and computational linguistics (*advancing methods for linguistic analysis and synthesis*).

8. Conclusion

This paper represents an effort to delineate the linguistic characteristics of Bengali cuisine texts, offering valuable insights into the unique features of this genre. By building the MBCC, the researchers have laid the groundwork for further exploration and application in various linguistic and computational fields. Future studies will continue to unravel the complexities of culinary language and its implications for broader linguistic and technological advancements.

This study is perhaps the first of its kind in any of the Indian languages. It furnishes authentic empirical evidence delineating the characteristic disparities between culinary language and other textual genres. Furthermore, it offers opportunities to explore the presence of many unique linguistic properties and patterns, the analysis of which sheds new light on the form and identity of the modern Bengali language. Leveraging the MBCC, there is also a plan to develop a Bengali cuisine lexical database and annotate it with multi-layered metadata (as proposed in this paper) for applications in machine learning, text processing, developing automated text-to-speech systems, and augmenting computational capabilities in the analysis and interpretation of written modern Bengali texts.

References

Ahmad, S. H., Omair, A., & Yousef, H. (2023). Creating and experimenting with a new parallel corpus of English-Arabic subtitles of culinary shows: a useful guide for translating food across cultures. *World Journal of English Language, 13*(8), 358-372.

Ashari, A. H., & Nugraha, S. I. (2022). Exploring culinary arts vocational high school students' learning of English specialised vocabulary. *Cendekia, 16*(2), 1-8.

Barlow, M. (2011). Corpus linguistics and theoretical linguistics. *International Journal of Corpus Linguistics, 16*(1), 3-44. (https://doi.org/10.1075/ijcl.16.1.02bar)

Chaki, J. B. (1996). *Bangla Bhasar Byakaran* (Grammar of the Bengali Language). Kolkata: Ananda Publishers.

Dash, N. S. (2005). *Corpus Linguistics and Language Technology: With Reference to Indian Languages*. New Delhi: Mittal Publications.

Dash, N. S. (2017). Understanding the form and composition of Bengali verbs as decontextualized autonomous lexical units. *Journal of Advanced Linguistics Studies, 6*(1&2), 158-188.

Dash, N. S., Dutta Chowdhury, P., & Sarkar, A. (2009). Naturalization of English words in modern Bengali: a corpus-based empirical study. *Language Forum, 35*(2), 127-142.

Dash, N. S., Bhattacharjee, M., & Dash, S. (2020). Foods offered to deities in Bengal: tasting ethnographic flavours in the gastronomic offerings prevalent in Bengali culture. *International Journal of Communication. 30*(1-2), 117-154.

Gerhardt, C., Frobenius, M., & Ley, S. (2013). *Culinary Linguistics: The Chef's Special*. Amsterdam: John Benjamins.

Heidari, T. H., & Jamalzadeh, M. (2020). Academic vocabulary in tourism research articles: A corpus-based study. *Journal of Language and Discourse Practice*, *1*(2), 25-44.

Jurafsky, D. (2014). *The Language of Food: A Linguist Reads the Menu*. New York: W.W. Norton & Company.

Laosrirattanachai, P., & Ruangjaroon, S. (2020). The word lists of hospitality service review construction. *Journal of Studies in the English Language, 15*(1), 107-158.

Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 50-86.

Lehrer, A. (1972). Cooking vocabulary and the culinary triangle of Levi-Strauss. *Anthropological Linguistics*, *14*(5), 155-171.

Lim, L. (2018). A corpus-based study of *braised* dishes in Chinese-English menus. Presented at the 32nd Pacific Asia Conference on Language, Information and Computation, the 25th Joint Workshop on Linguistics and Language Processing, Hong Kong, 1-3 December 2018.

Mallick, B. P., Bhattacharya, N., Kundu, S. C., & Dawn, M. (1998). *Phonemic and Morphemic Frequency in the Bengali Language.* Kolkata: Asiatic Society.

Mathew, E. (2024). Globalization and local flavours: the impact of modern food production on traditional cuisine and culinary heritage preservation. *International Journal for Multidimensional Research Perspectives, 2*(7), 61-74.

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

Muysken, P. (1997). Code-switching processes: alternation, insertion, and congruent lexicalization. *Language Choices: Conditions, Constraints, and Consequences*, *20*, 361-380.

Myers-Scotton, C., & Jake, J. L. (2001). Explaining aspects of codeswitching and their implications. In: Nicole, J. (Ed.) *One Mind, Two Languages: Bilingual Language Processing* (pp. 84-116). Oxford:

Blackwell.

Myers-Scotton, C. (1983). The negotiation of identities in conversation: a theory of markedness and code choice. *International Journal of the Sociology of Language. 44,* 115-136.

Nandy, A. (2002). Ethnic cuisine: the significant 'other'. *India International Centre Quarterly, 29*(3&4), 246-251.

Nordin, M. N. R., Stapa, S. H., & Darus, S. (2013). Developing a specialized vocabulary word list in a composition culinary course through lecture notes. *Advances in Language and Literary Studies, 4*(1), 78-88.

Popovski, G., Seljak, B. K., & Eftimov, T. (2019). FoodBase corpus: a new resource of annotated food entities. *Database, e-version,* 1-13. doi: 10.1093/database/baz121

Riyantika, F., Endra, R. Y., Afrulli, T., & Puspa, A. K. (2020). Needs analysis of English for specific purposes at food and beverage Emersia hotel in Bandar Lampung. *Beyond Linguistika (Journal of Linguistics and Language Education), 3*(2), 7-11.

Rungrueang, T., Pitchapha, B., Sitanan, P., & Piyapong L. (2022). Corpus-based approach to generate a word list for food service. *Thai TESOL Journal. 35*(1), 57-76.

Sengupta, J. (2010). Nation on a platter: The culture and politics of food and cuisine in colonial Bengal. *Modern Asian Studies, 44*(1), 81-98.

Thompson, H.-R. (2010). *Bengali: A Comprehensive Grammar.* London and New York: Routledge.

Xi, C., & Xie, W. (2024). Unveiling hybrid culinary cultures in Macao: a corpus-assisted multimodal approach to the translation of Macanese cuisine. *Perspectives, 1,* 1-19.

Xu, W. (2007). *Eating Identities: Reading Food in Asian American Literature.* Hawaii: University of Hawai'i Press.

Zhang, M. (2009). A corpus-based comparative study of semi-technical and technical vocabulary. *The Asian EFL Journal, 9*(2), 148-172.

Zhuoma, Q. (2023). A corpus analysis on food-related borrowing words between English and Chinese and its influences on cultural implication and language development. *British Journal of English Linguistics, 11*(3), 33-45.

THE AUTHORS

Professor Niladri Sekhar Dash works in the areas of corpus linguistics, language processing, computational lexicography, indigenous language documentation & digitization, computer-assisted language teaching, and digital humanities and ethnography. He has published 24 books/monographs and more than 220 research papers. He has delivered academic talks and taught courses in more than 150 universities and institutions in India and abroad.

Dr. Amrita Bhattacharyya is an Associate Professor at Amity Institute of English Studies and Research, Amity University Kolkata. She is a bilingual poet and a writer having thirteen books published to her credit. Her areas of research interest include Culture Studies, Metaphysics, and Linguistics. She has published several research papers and book chapters in national and international peer-reviewed journals and books, e.g., Routledge, Palgrave Macmillan, IEEE journal, Rupkatha Journal, JCLA, CIIL Mysuru, Sanglap, to name a few.

THE AUTHORS' ADDRESSES

First Author
Niladri Sekhar Dash
Professor
Indian Statistical Institute, Kolkata
Barrackpore Trunk Road Kolkata, INDIA
Email: ns_dash@yahoo.com

Corresponding Author
Amrita Bhattacharyya
Associate Professor
Amity University, Kolkata
Major Arterial Road, Action Area II, Rajarhat, New Town, Kolkata, INDIA
Email: arushi.asmi13@gmail.com