

# Natural Language Processing of Spoken Diet Records (SDRs)

Ronilda Lacson MD, PhD<sup>a,b</sup> and William Long PhD<sup>a</sup>

Massachusetts Institute of Technology (MIT), Cambridge, MA<sup>a</sup>

Decision Systems Group, Brigham and Women's Hospital, Boston, MA<sup>b</sup>

## Abstract

*Dietary assessment is a fundamental aspect of nutritional evaluation that is essential for management of obesity as well as for assessing dietary impact on chronic diseases. Various methods have been used for dietary assessment including written records, 24-hour recalls, and food frequency questionnaires. The use of mobile phones to provide real-time dietary records provides potential advantages for accessibility, ease of use and automated documentation. However, understanding even a perfect transcript of spoken dietary records (SDRs) is challenging for people. This work presents a first step towards automatic analysis of SDRs. Our approach consists of four steps – identification of food items, identification of food quantifiers, classification of food quantifiers and temporal annotation. Our method enables automatic extraction of dietary information from SDRs, which in turn allows automated mapping to a readily available dietary database. Our model has an accuracy of 90%. This work demonstrates the feasibility of automatically processing SDRs.*

## I. Introduction:

Dietary assessment is a fundamental aspect of nutritional evaluation that is essential for the diagnosis and monitoring of food-based disorders such as inadequate protein or calorie nutrition, inappropriate food choices, or management of over-eating and obesity. Furthermore, diet plays a crucial role in the etiology and management of chronic diseases such as kidney diseases, heart diseases, diabetes and cancer. The most commonly used methods for dietary assessment include (1) diet records, in which respondents document all food items consumed over a given period of time, usually for a period of three and seven days; (2) 24-hour recalls where respondents are asked to remember all food items they ingested over the past 24 hours, usually repeated for a period of three to seven days; (3) diet history, which involves questioning respondents regarding their "typical" or "usual" food intake in a "prolonged" interview in order to construct a typical seven days' eating pattern.; and (4) food frequency questionnaires (FFQ), in which respondents are presented with a long list of food items and asked to report their estimated frequency of consumption for each food item over a specific time period, from periods as short as one week up to

periods as long as one year. The first three approaches are deemed to be more accurate and timely but require substantial subject and investigator time, effort, and expense. Often, these records may still be incomplete as a result of human error and lapses.<sup>1,2</sup> The questionnaires have become popular for documenting absolute dietary intake and/or dietary adherence in clinical trials because of ease of pre-coding and standardized entry into a database.<sup>3,4</sup> However, despite increased utility for describing the relationship between diet and disease in population studies, the error characteristics of the various FFQs preclude an accurate assessment of individual intake and render the process fastidious, at best.<sup>2,5</sup> For instance, in determining links between diet and cancer development, food records have established plausible relationships which have not been elicited with the application of FFQs.<sup>6</sup> There is no question that diet diaries completed in "real-time" provide more specific information and detail (e.g. food preparation such as frying vs. broiling) than is obtainable from FFQs.<sup>7</sup> The availability of information such as food preparation, meal combinations, mixed dishes, food type, brand name, portion size, among other details, provides maximum flexibility for investigators in the conduct of analyzing nutritional intake and dietary choices. However, the classic methodology requires a writing instrument and food diary to always be close at hand. As such, it can be very tedious for the patient to complete the record and it is not uncommon for the diary to be conveniently unavailable during mealtimes, especially when eating outside the home. We believe that we have identified a more convenient tool and developed a more efficient procedure for collecting such valuable information.

The proposed utility of mobile phones to provide a real-time dietary record provides potential advantages for accessibility, ease of use, and options for near-instantaneous transfer of information and automated documentation. Mobile phones are increasingly becoming ubiquitous in the general population and tend to be within a person's reach at most times. Apart from the spoken dietary record (SDR) that forms the first phase of this revolutionary approach, there is the potential for utilizing new developments in technology such as photo or video capability for sending pictures or short video recordings of the food and beverage intake, both before and after the meal. The rapid evolution of mobile phone technology

coupled with continued advancement in automatic image recognition provides an additional layer of information that may assist researchers in more precise estimation of true dietary intake. Therefore, harnessing the mobile phone provides a revolutionary, alternative method to dietary assessment.<sup>8</sup>

Despite the overall vision for developing this tool, the project has to begin with a critical basic component of this new modality – automatic processing of SDRs. To illustrate the difficulty in manually accessing recorded data – even data that has been manually transcribed – consider this example of 30 seconds of an error-free transcript (see Figure 1). Similar to problems encountered in analyzing medical dialogues, we note that SDRs exhibit an informal, verbose style, characterized by interleaved false starts (such as “I... at three thirty”) and non-lexical filled pauses (such as “ah” and “umm”).<sup>9</sup> This exposition also highlights the striking lack of structure in the transcript: a description of food items ingested that day switches to a description of the previous evening’s snack without any visible delineation customary in written text. Such disorganized verbal descriptions may prove difficult to discern or manually record, leading to errors and variability in documentation.

Figure 1: Transcribed segment of an SDR

<p>I at three thirty we were at a reception after a funeral and I had a small sandwich umm ham and cheese it was a half slice of bread size ah and two cookies and ah a serving of fruit pineapple and strawberries and ah a glass of juice and a cup of coffee last night we had a late evening snack I ah had a piece of bread with jam and cheese on it a slice of jarlsberg cheese on the piece of bread and I had a salad also and that was it</p>
---

The goals of this paper are two-fold: (1) We propose a framework for acquiring the structure of SDRs; and (2) We present and evaluate a four-step algorithm that enables automatic extraction of dietary information from these records, which in turn allows automated mapping to the Diet History Questionnaire dietary/nutrient database, a commonly used dietary assessment instrument based on 4,200 individual foods reported by adults in the 1994-1996 US Department of Agriculture Continuing Survey of Food Intakes by Individuals (CSFII).<sup>10</sup>

**II. Materials and Method:**

**Data Collection**

We recorded SDRs from six individuals for seven to twenty days each. The only instructions given include

(1) recording all food and drinks ingested during the specified time period; (2) including applicable food portions (e.g. a cup of coffee, two fried chicken thighs); and (3) for food items that are complex or unusual in their own judgment, we asked that ingredients be specified, if possible. There was no prior training and there were no restrictions in word usage or timing of the recordings, although recording as close as possible to the meals was encouraged.

We used a Nokia 6600 mobile phone as a voice recording device to record descriptions of dietary information. Each SDR is automatically stored in the mobile phone with its corresponding date/time stamp. All SDRs were downloaded to a computer after twenty days and manually transcribed by the investigator (RL), maintaining delineations between individual recordings.

The data were then divided into training and testing sets according to the chronological order in which they were received. The data characteristics for each set are shown in Table I.

Table 1: Data characteristics

	Training Set	Test Set
Number of recordings	86	42
Number of words	3812	1863
Number of meals	117	61
Number of food items	516	289

Figure 2: Examples of Annotated Food Items

Actual Transcript of SDR	for breakfast this morning i had a bowl of <cereal> with <blueberries> on top and <milk> and a small <orange juice> and a cup of <coffee>
Annotated SDR	for breakfast this morning i had a bowl of <food> with <food> on top and <food> and a small <food> and a cup of <food>

**Data Processing**

The automatic processing of the data is divided into four steps.

**Step 1:** Identification of Food Items – Our goal is to identify all food items in the SDRs. Our approach builds on the food items that are available in our food database, the Diet History Questionnaire dietary/nutrient database. In addition, all food items in the training set were manually identified and matched to food items in the database. Each term in the data that matches a term in the augmented database is automatically labeled, with multiple word

matches given preference over single word matches (e.g. “orange juice” is a better match than “orange” and “juice”). An example of this annotation is shown in Figure 2.

**Step 2:** Quantification of Food Items – This step utilizes supervised classification to identify appropriate terms used to quantify food items. The basic model we present here follows the traditional design of a word classifier similar to those used for word sense disambiguation.<sup>11</sup> It predicts the semantic type of a word based on a shallow meaning representation encoded as simple lexical features. In our case, we predict whether each word is a food quantifier or not.

We allowed unlimited word usage for the SDRs and we noted a wide diversity in how food quantities were recorded. For example, chicken was reported with the following amounts: “four ounces,” “two drumsticks,” “a small portion,” “a thigh” or “four pieces.” Each word in the training and test sets was labeled manually by one of the investigators.(RL) We describe below the framework for the supervised classification that we used, including feature selection and combination.

**Feature selection.** Our model for two-way classification relies on three features that can be easily extracted from the transcript: the actual words, the two preceding words and the two subsequent words. Clearly, words are highly predictive of their class. We expect that food quantifiers would contain words like “one,” “small,” “cup” and “tablespoon.” The preceding words are also likely to improve classification. Words that typically precede food quantifiers include “I had”, “I ate” and “and”. It is likewise important to include subsequent words in our list of features. Most food quantifiers are used to describe the food and would typically occur before a food item. Thus, words such as “of <food>” or “<food>” follow food quantifiers. For example in Table 1, in the phrase “I had a bowl of cereal”, “a bowl” is a food quantifier that occurs after “I had” and before “of <food>”.

**Feature weighting and combination:** We used Boostexter,<sup>12</sup> a boosting classifier, to perform supervised classification. Each word in the training set is represented as a vector of features and its corresponding class. Using the training set, boosting works by initially learning simple weighted rules, each one using a single feature to predict one of the labels with some weight. It then searches greedily for the subset of features that predict a label with high accuracy. Using the test data set, the algorithm

outputs the label with the highest weighted vote to classify each word.

**Step 3:** Classification of Food Quantifiers – The Diet History Questionnaire dietary/nutrient database divides food portions into three sizes, which we arbitrarily labeled small, medium and large. The food portions obviously depend on the food being quantified and how the food is measured (e.g. “a cup”, “one tablespoon”, “two pieces”). A cup of ravioli, for example, would be considered medium, whereas a cup of “maple syrup” would be considered large. Thus, we used supervised classification, similar to the method we employed in Step 2, in order to classify the food quantifiers into small, medium and large.

**Feature selection and combination.** Our model for three-way classification relies only on two features: the actual words used as food quantifiers and the food item it describes. The food item associated with a food quantifier is automatically computed by selecting the one that occurs closest to the food quantifier by counting the number of words between a food item and the food quantifier. For example, in the phrase “a bowl of cereal with blueberries”, the food quantifier “a bowl” would describe “cereal” instead of “blueberries” because only one word separates it from “cereal” (“of”) whereas three words separate it from “blueberries” (“of cereal with”). Food items can only be associated with a single food quantifier. To break ties, a quantifier is associated with the food item subsequent to it. In the example “a small orange juice and a cup of coffee”, “a small” refers to “orange juice” because of proximity. The quantifier “a cup” refers appropriately to “coffee” because “orange juice” already has a quantifier associated with it. Each food quantifier in the training set is represented as a vector of features (the food quantifier and the associated food item) and its corresponding class (small, medium or large). Boostexter is used to perform supervised classification.

**Step 4:** Temporal Annotation – The mobile phone automatically records the exact date and time when the SDRs were recorded. This, however, is not sufficient to appropriately label when a food item was ingested. For example, “I ate an apple for breakfast yesterday” should include “apple” among the food items ingested on the previous date and not the date when the SDR was recorded. We used GUTime time expression tagger to perform the appropriate date/time annotations.<sup>13</sup> GUTime is a program that tags and normalizes time expressions in natural language text. It first recognizes time expressions from part-of-speech-tagged text. The

time expression can be absolute (e.g. Feb. 14, 2005) or relative (e.g. Tuesday). It then uses variables including verb tense, nearby dates and explicit offsets (e.g. “next” Tuesday) to assign actual dates to relative time expressions. Using the example in Figure 1, we obtain the text with appropriate date tags as shown in Figure 3. Thus, we are able to accurately label the date when food items were ingested.

Figure 3: Date-tagged SDRs

```
<TIMEX3 VAL="20051223">12/23/2005</TIMEX3>
I at three thirty we were at a reception after a
funeral and I had a small sandwich umm ham and
cheese it was a half slice of bread size ah and two
cookies and ah a serving of fruit pineapple and
strawberries and ah a glass of juice and a cup of
coffee
<TIMEX3 VAL="20051222">last night</TIMEX3>
we had a late evening snack I ah had a piece of
bread with jam and cheese on it a slice of jarlsberg
cheese on the piece of bread and I had a salad also
and that was it
```

### III. Results:

Table 2 shows the accuracy of the algorithms at each step described above. The baseline model is given by classifying each item with the most frequent class. In Steps 1 and 2, this would mean labeling all words as not being food items or not being food quantifiers, respectively. In Step 3, this would label all food quantifiers as having medium portions. Finally, in Step 4, this would label all dates for all meals as being the date of recording the SDR. Proceeding step-wise, we obtain a combined accuracy of 0.90 in identifying all food items with their appropriate food quantifiers with accurate temporal annotation in the test data set.

Table 2: Models and their accuracy

Model	Accuracy	Baseline	p-value (Fisher’s Exact)
Identification of Food Items	0.95	0.85	p<0.0001
Quantification of Food Items	0.98	0.84	p<0.0001
Classification of Food Quantifiers	0.92	0.74	p<0.0001
Temporal Annotation	1.00	0.54	p<0.0001

We show that at each step, the methods we describe for automatically processing SDRs significantly outperform baseline. In fact, the methods were reasonably successful in automatically extracting all

relevant nutritional information from transcribed unstructured SDRs. As expected, the food items were easily identifiable, given the food database and the prior food items that are reported in the training data. Likewise, the classifier we developed for identifying food quantifiers using the lexical features we identified performed accurately. Examples of predictive features for food quantifiers are shown in Table 3. It appears that relying solely on words in close proximity to the current term can accurately predict whether the term is a food quantifier or not. As noted in Part II (Step 2), “of” and “food” are in fact significant predictors of a term being a food quantifier if “of” and/or “food” occur after the given term (e.g. “a bowl of food” where “a bowl” is a food quantifier).

Table 3: Examples of predictive features for identifying food quantifiers

Category	Current word	Preceding word	Subsequent word
Food Quantifier	half, cups, piece, one, small	also, with, food	of, food, serving, amount
Not a Food Quantifier	food, and, of	of, ounces	had, four, the

In step 3, we are able to perform classification of food quantities into three distinct portions reasonably well given the food items and their associated food quantifiers. We show examples of the most predictive features for classifying food quantifiers in Table 4. As expected, both food items and quantifiers play a role in predicting whether food portions are classified as small, medium or large.

Table 4: Examples of predictive features for classifying food quantifiers into three categories

Category	Food Item	Food Quantifier
Small	toast, wine, soup	small, slice, few, fourth
Medium	carrots, porridge	medium, slice
Large	applesauce	ten, large

Finally, the time tagger performed with perfect accuracy on the test data set. It was able to correctly identify the dates when each meal was ingested.

### IV. Discussion:

We describe a method for automatically processing manually transcribed SDRs. We determine that a spoken dietary record consists of three key elements – food items, food quantifiers, and the date when the food is ingested. These key elements enable mapping an unstructured SDR into a structured and widely used dietary/nutrient database. Although we do not

have perfect accuracy, the current system is able to fully automate the process of analyzing SDRs with reasonable accuracy (90%). Perfect accuracy in this evaluation assumes that all food items are identified with their corresponding quantity and classified into the three portions that were utilized in the dietary database. We note that most of the loss in accuracy stems from this food quantifier classification (small, medium, large), which may be arbitrary. Using actual food quantifiers (e.g. “one cup of broccoli”) in processing SDRs is another option that this system can provide. Exchange units can be used to quantify the amounts of calories and nutrients in food items.

We present a fully-implemented stepwise algorithm for automatically extracting key dietary information. We believe this is a significant and innovative approach to dietary assessment that has minimal burden to both consumer and health care provider. The use of natural language enhances the usability of the system and the wide availability of mobile phones makes it an ideal medium for collecting this information. The current algorithm focuses on natural language processing techniques that enable automatic processing of the SDRs. Our emphasis on spoken records sets us apart from efforts to interpret written medical text.<sup>14,15</sup>

We plan to extend this work in the future in four major directions. First, we plan to apply our method to more SDRs in order to address the limited variability in vocabulary and structure of the SDRs. Second, we plan to apply our method to automatically recognized conversations. To maintain the classification accuracy, we will explore the use of acoustic features to compensate for recognition errors in the transcript. Third, we will explore other databases that allow more granularities in representing food quantifiers. Using more specific food quantifiers might provide better accuracy in quantifying dietary intake. Lastly, we plan to take photographs of the actual food items. Images may help provide more accurate records of food intake and food portions. In addition, images may improve patients’ recollection of food that they ingested in previous meals. Automatic image recognition and speech recognition promise to extend the capabilities of this new modality in dietary assessment.

#### Acknowledgements

*The authors would like to thank Dr. Eduardo Lacson, Dr. Peter Szolovits and the anonymous reviewers from AMIA for their helpful comments and suggestions.*

#### V. References:

1. Klesges R, Eck L, Ray J. Who underreports dietary intake in a dietary recall? Evidence from the

- Second National Health and Nutrition Examination Survey. *J Consult Clin Psychol.* 1995; 73: 438-44.
2. Schaefer E, Augustin J, Schaefer M, et al. Lack of efficacy of a food-frequency questionnaire in assessing dietary macronutrient intakes in subjects consuming diets of known composition. *Am J Clin Nutr.* 2000; 71(3): 746-51.
3. Zulkifli SN, Yu SM. The food frequency method for dietary assessment. *J Am Diet Assoc.* 1992 Jun;92(6):681-5.
4. Thompson FE, Moler JE, Freedman LS, Clifford CK, Stables GJ, Willett WC. Register of dietary assessment calibration-validation studies: a status report. *Am J Clin Nutr.* 1997 Apr;65(4 Suppl):1142S-1147S.
5. Liu K. Statistical issues related to semiquantitative food-frequency questionnaires. *Am J Clin Nutr.* 1994 Jan;59(1 Suppl):262S-265S.
6. Bingham S, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet.* 2003; 362:212-4.
7. Sempos CT, Liu K, Ernst ND. Food and nutrient exposures: what to consider when evaluating epidemiologic evidence. *Am J Clin Nutr.* 1999 Jun; 69(6):1330S-1338S.
8. Kristal A, Peters U, Potter J. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev.* 2005; 14(12):2826-8.
9. Lacson R, Barzilay R. Automatic processing of spoken dialogue in the home hemodialysis domain. *Proc. AMIA Symposium,* 2005.
10. Flood A, Subar A, Hull S, Zimmerman T, Jenkins D, Schatzkin A. Methodology for adding glycemic load values to the National Cancer Institute diet history questionnaire database. *Journal of the American Dietetic Association.* 2006; 106(3): 393-402.
11. Liu H, Teller V, Friedman C. A Multi-aspect Comparison Study of Supervised Word Sense Disambiguation. *J Am Med Inform Assoc.* 2004 Jul-Aug; 11(4): 320-331.
12. Schapire R, Singer Y. Boostexter: A boosting-based system for text categorization. *Machine Learning.* 2000; 39(2/3):135-168.
13. Mani I, Wilson G. Processing of news. *Proc. 38th Annual Meeting of the Association for Computational Linguistics.* 2000: 69-76.
14. Xu H, Anderson K, Grann VR, Friedman C. Facilitating cancer research using natural language processing of pathology reports. *Medinfo.* 2004:565-72.
15. Hsieh Y, Hardardottir GA, Brennan PF. Linguistic analysis: Terms and phrases used by patients in e-mail messages to nurses. *Medinfo.* 2004:511-5.