DATA COLLECTION AND LANGUAGE UNDERSTANDING OF FOOD DESCRIPTIONS

M. Korpusik, N. Schmidt, J. Drexler, S. Cyphers, J. Glass
MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, U.S.A.

Introduction

- Existing approaches for the prevention and treatment of obesity are hampered by the lack of accurate, low-burden methods for self-assessment of food intake, especially for hard-to-reach, low-literate populations.
- ***Goal****: create a nutrition dialogue system that automatically extracts foods from a user's spoken meal log.*
  - E.g., "This morning I ate **a bowl of Kellogg's cereal**."
- We have explored two components:
  - Data collection
  - Language understanding (i.e., tagging and segmenting food concepts)

Methods

Tagging

- ***Goal****: label foods and properties in a spoken meal description.*

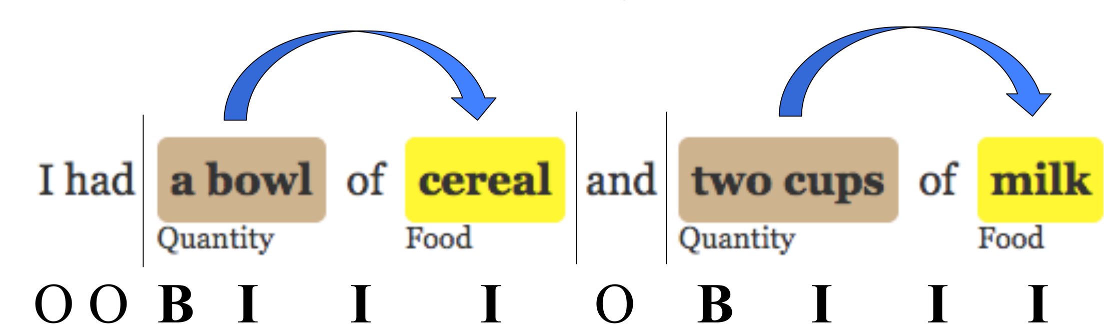
- Semi-Markov conditional random field (semi-CRF): outputs labels for *segments* of input tokens.
- Chose features using 10-fold cross-validation: n-grams, food lexicon, and POS tags.

Features	Mean F1	Variance	St. Dev.
N-grams	84.6	0.6	0.8
+ Lexicon	84.6	1.3	1.1
+ POS tags	**84.8**	1.2	1.1

Table 1. Semi-CRF 10-fold results.

Segmenting

- ***Goal****: associate property labels with food labels.*
- Framed as a BIO labeling classification task.



- Simple rule baseline: assign properties to subsequent foods.
- Investigated three approaches:
  1. CRF classifier (CRF++ toolkit)
  2. Markov model (MM)
     - Each state represents a property or food.
     - Used a finite state transducer (FST).
  3. Transformation-based learning (TBL)
     - Iteratively applies transformations to improve an initial solution.

Segmenting Results

- TBL improved upon the simple rule and Markov model baselines.
- CRF significantly better than other methods.

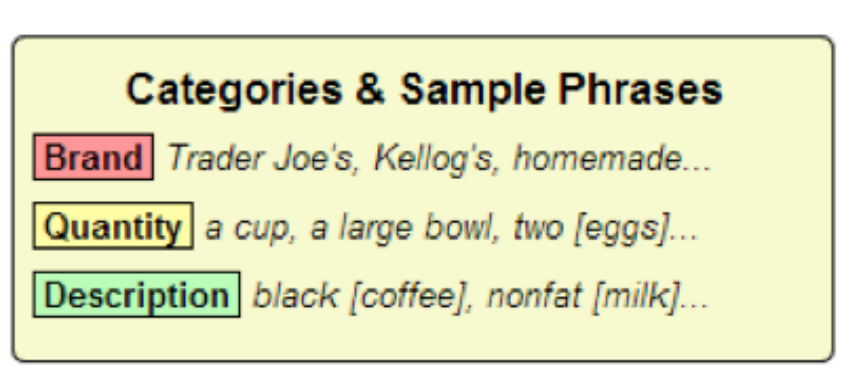
Approach	Acc	Prec	Recall	F1
Simple Rule	84.4	51.5	54.2	52.8
Simple + TBL	94.3	77.9	78.3	78.1
MM	84.9	54.6	57.2	55.9
MM + TBL	95.2	82.7	80.4	81.5
CRF	**97.2**	**87.1**	**87.1**	**87.1**
CRF + TBL	95.5	84.0	82.9	83.4

Table 3. Performance on food segmenting task (token-level accuracy and phrase-level F1).

- Simple rule incorrectly assigns properties if attribute comes after its corresponding food.
- Markov model makes mistakes by incorrectly segmenting foods.

Data Collection

- We collected and labeled 1,302 breakfast diaries on Amazon Mechanical Turk (AMT).
- Three rounds:
  - Writing meal descriptions
  - Labeling foods
  - Labeling properties (i.e., brand, quantity, description)



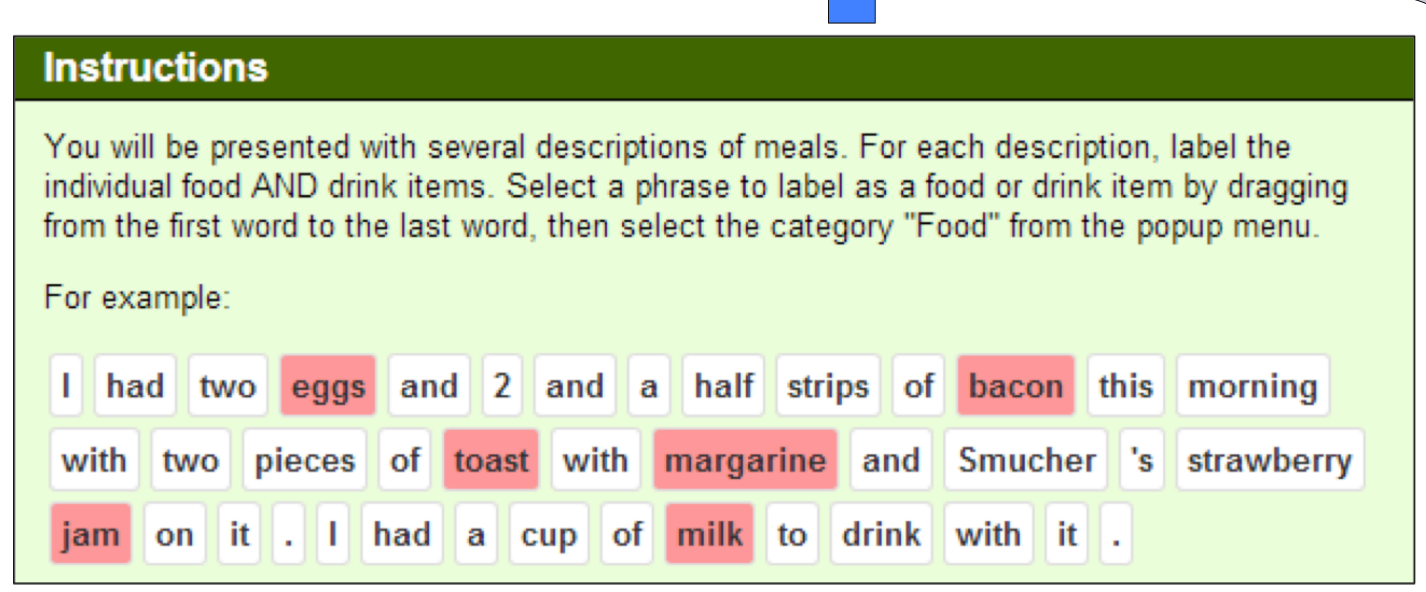


Fig. 1. The AMT task for labeling foods in a meal description.

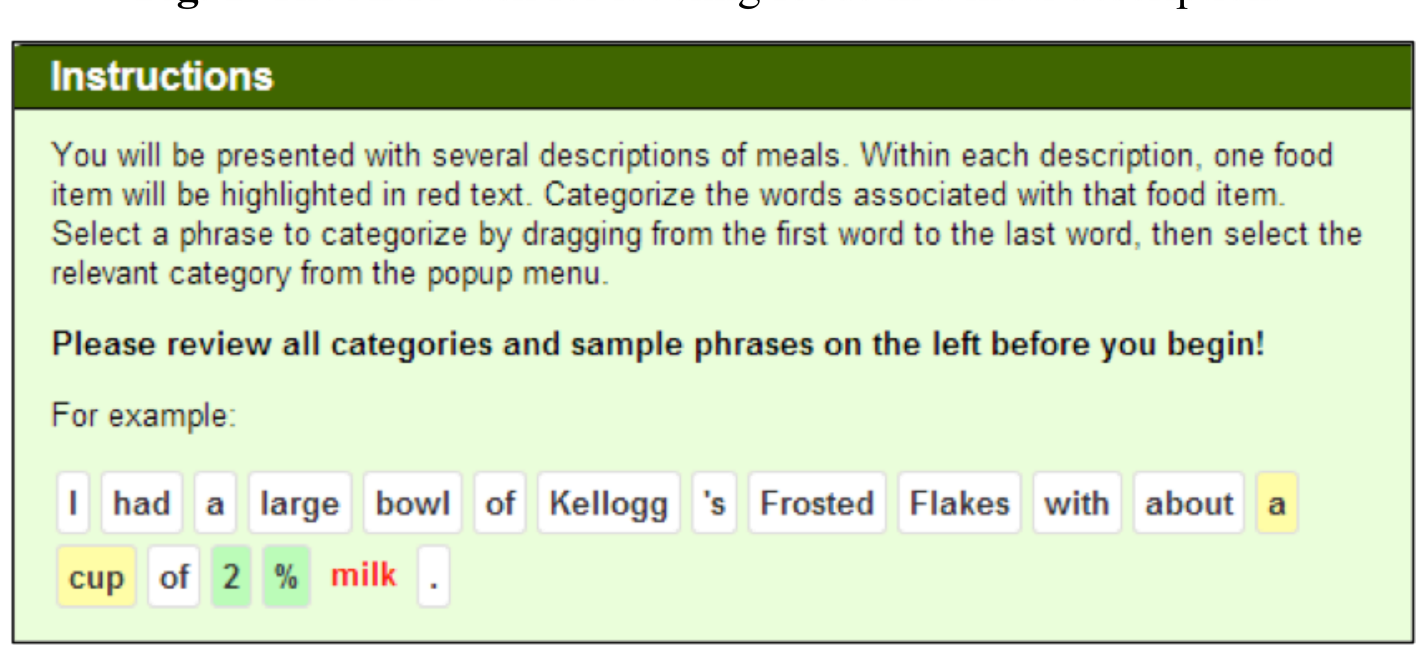


Fig. 2. The AMT task for labeling properties of foods.

Tagging Results

- Evaluated semi-CRF performance at the food concept level, rather than at the token level.

Label	Precision	Recall	F1
Food	92.5	87.5	89.9
Brand	87.3	71.0	78.3
Quantity	92.4	91.3	91.9
Description	85.6	77.6	81.4
Other	91.7	95.8	93.7
Overall	88.3	82.2	85.1

Table 2. Semi-CRF concept-level performance.

- Foods, quantities, and other were labeled more accurately than brands or descriptions.
- Semi-CRF performance was not significantly different from CRF (significance measured using McNemar's test).

Conclusions

- Performed data collection and annotation of food diaries via Amazon Mechanical Turk.
- Conducted semantic labeling experiments using a semi-CRF with an F1 test score of 85.1.
- Explored three methods for associating foods with their corresponding attributes: a Markov model (MM), transformation-based learning (TBL), and a CRF classifier.
- CRF is the best food segmenting model, achieving a phrase-level F1 score of 87.1.

Ongoing Work

- Asking follow-up questions to narrow down the database hits.
- Mapping user's spoken quantities to database quantities.
- Refining the user interface.