

Understanding User Reviewing Patterns: URSA

Gayatree Ganu¹, Amélie Marian¹, Noémie Elhadad²

¹Dept. of Computer Science, Rutgers University, NJ, USA; ²Dept. of Biomedical Informatics, Columbia University, NY, USA

Contact: gganu@cs.rutgers.edu

Motivation and Background

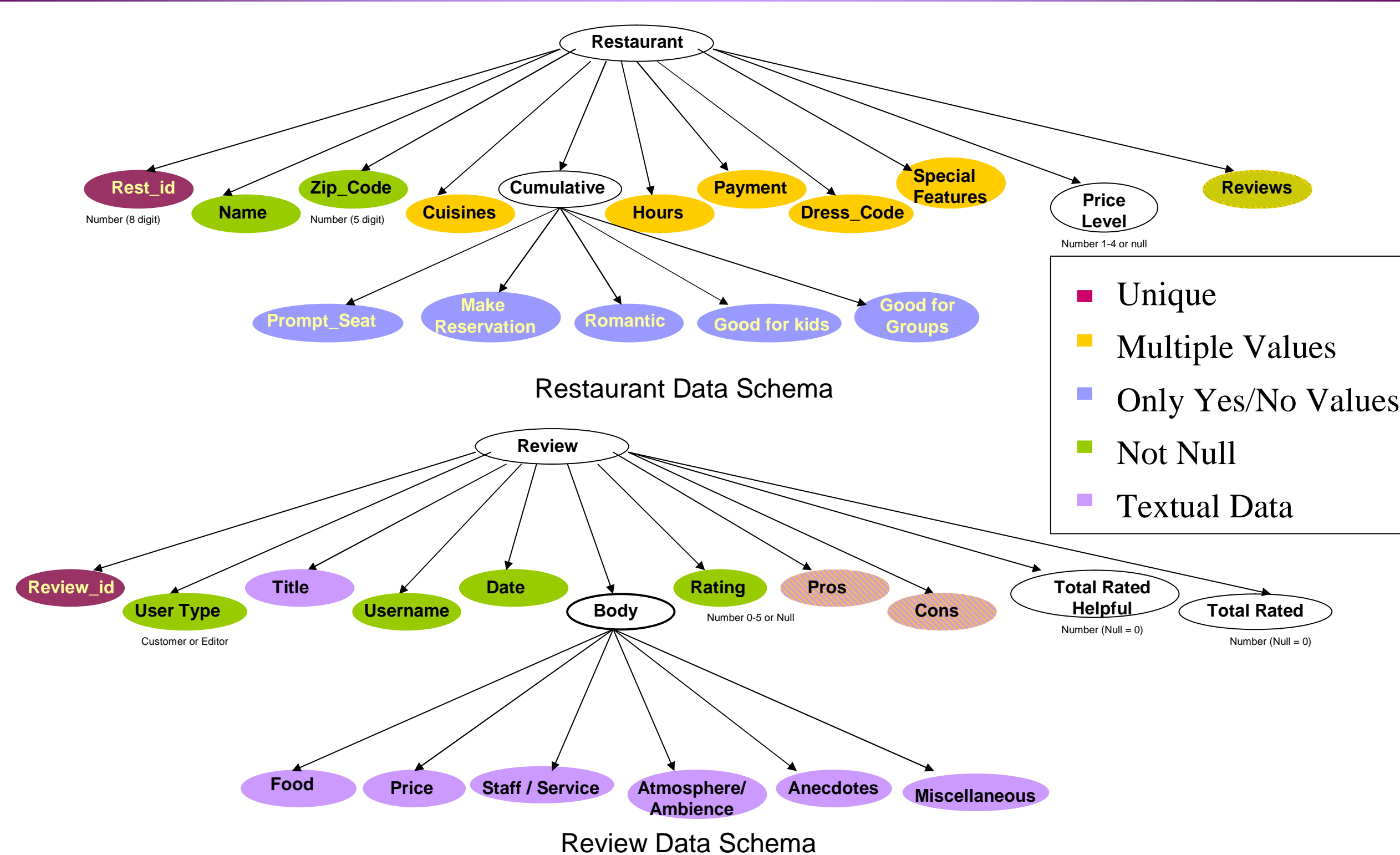
- Websites commonly allow users to input reviews.
- Most reviews are written in free-text format with very scant structured metadata information.
- Difficult to understand, analyze and aggregate textual reviews.

In the URSA (User Review Structure Analysis) project, we provide a better understanding of user review patterns and develop tools to better search, understand and access reviews.

Dataset : Restaurant Reviews

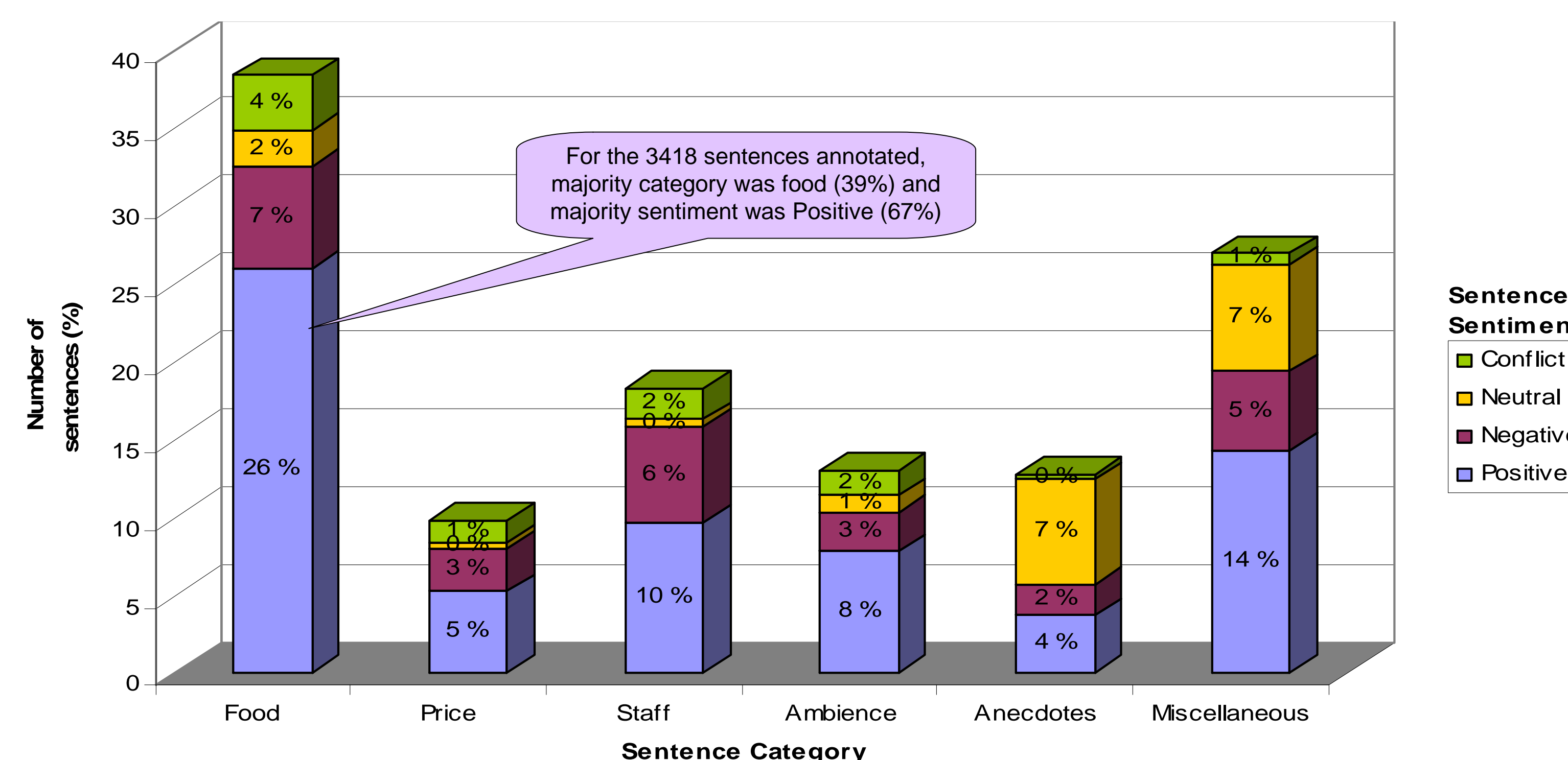
- Data is mined from <http://newyork.citysearch.com> (8/1/06 through 8/7/06).
- 5531 restaurants
- 50877 reviews
 - maximum number of reviews for a restaurant is 242
 - 28 restaurants have 100 or more reviews.
 - 1340 restaurants have 10 or more review.
- 32167 different users, with 375 users with more than 10 reviews.

Data Schema

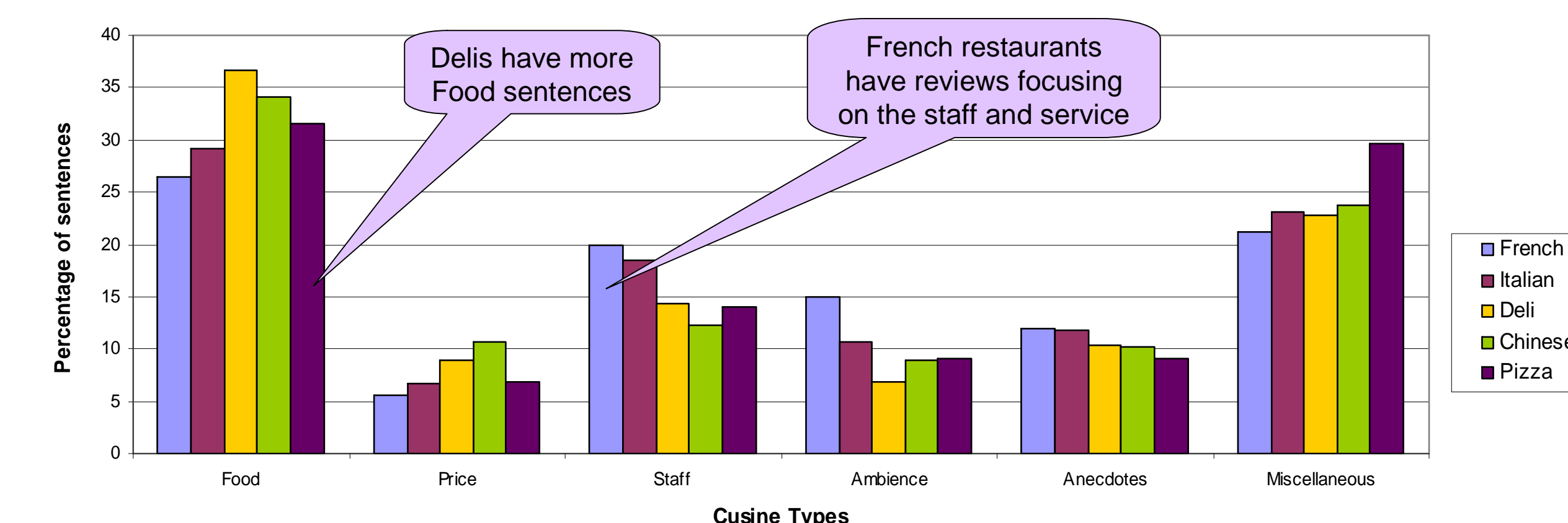


Preliminary Results

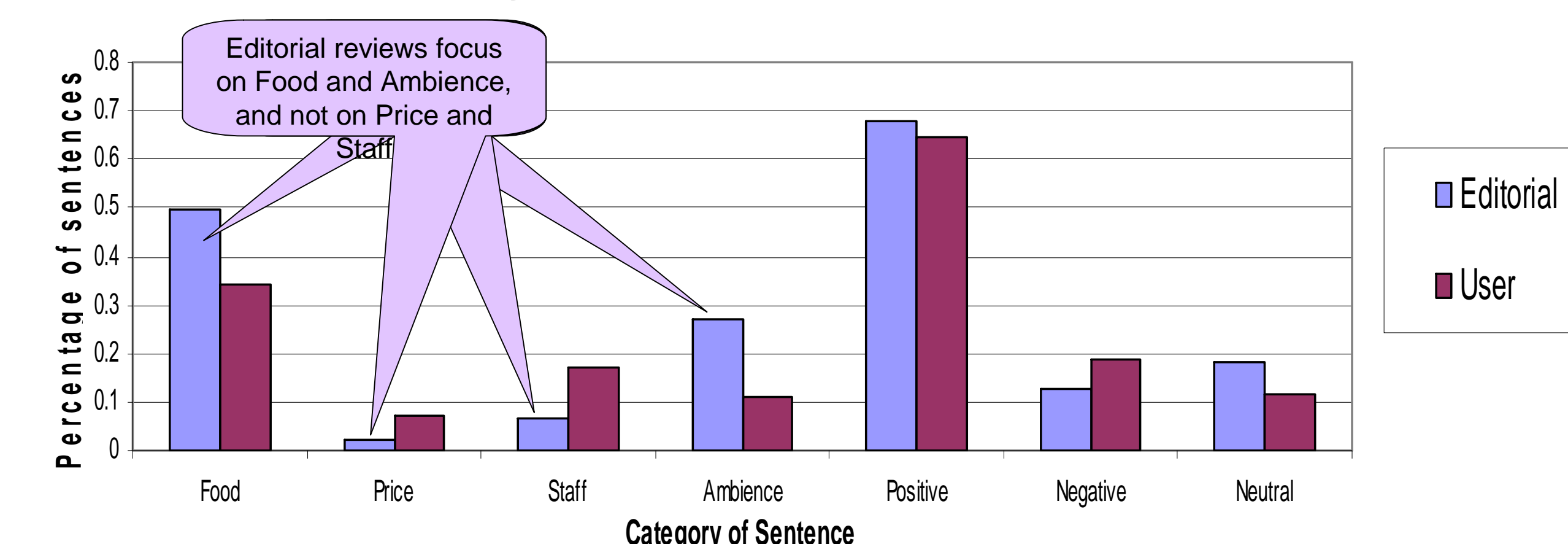
Sentence classification based on category and sentiment



Distribution of sentences based on Restaurant Cuisine

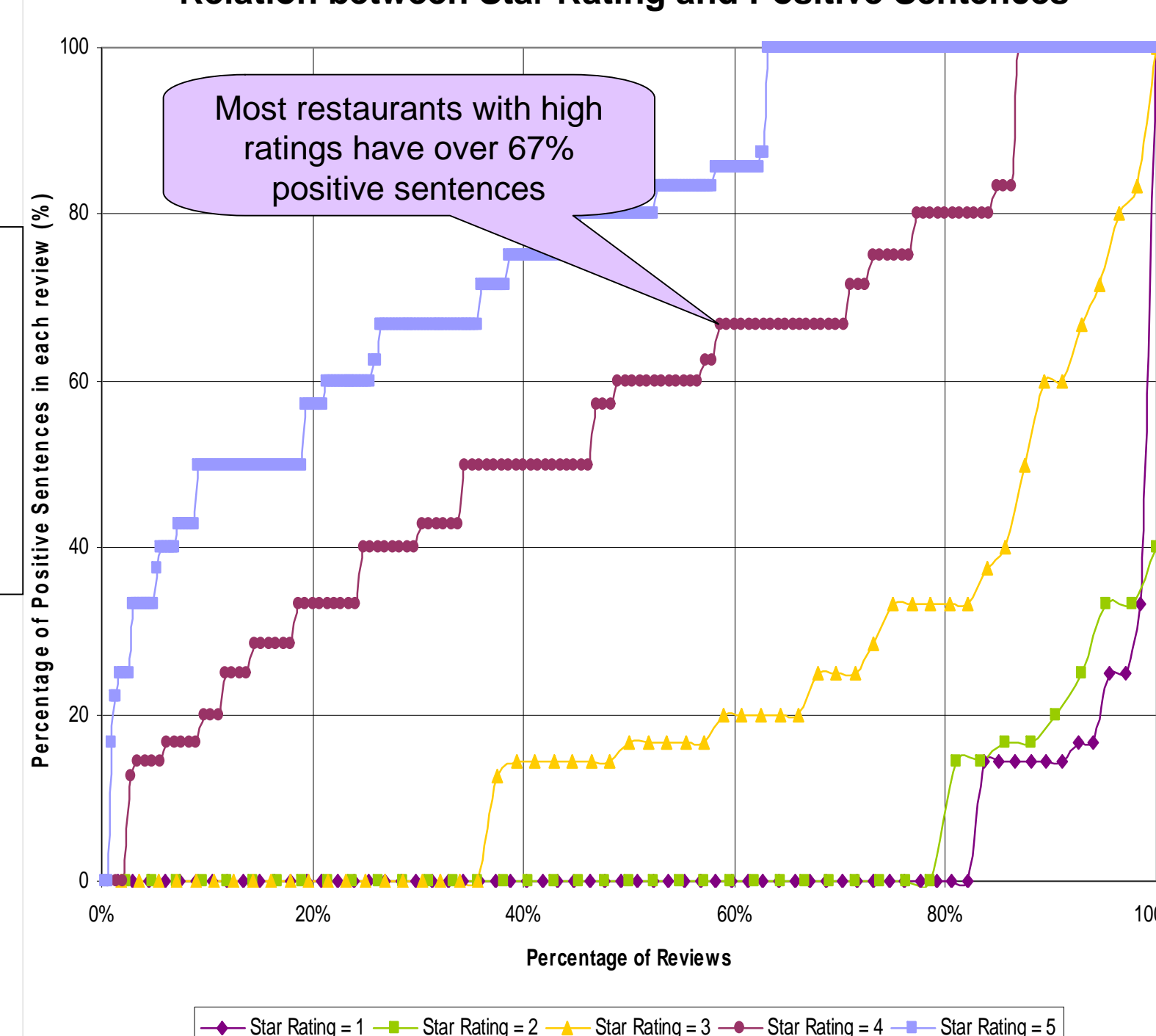


Comparing Editorial reviews with User reviews

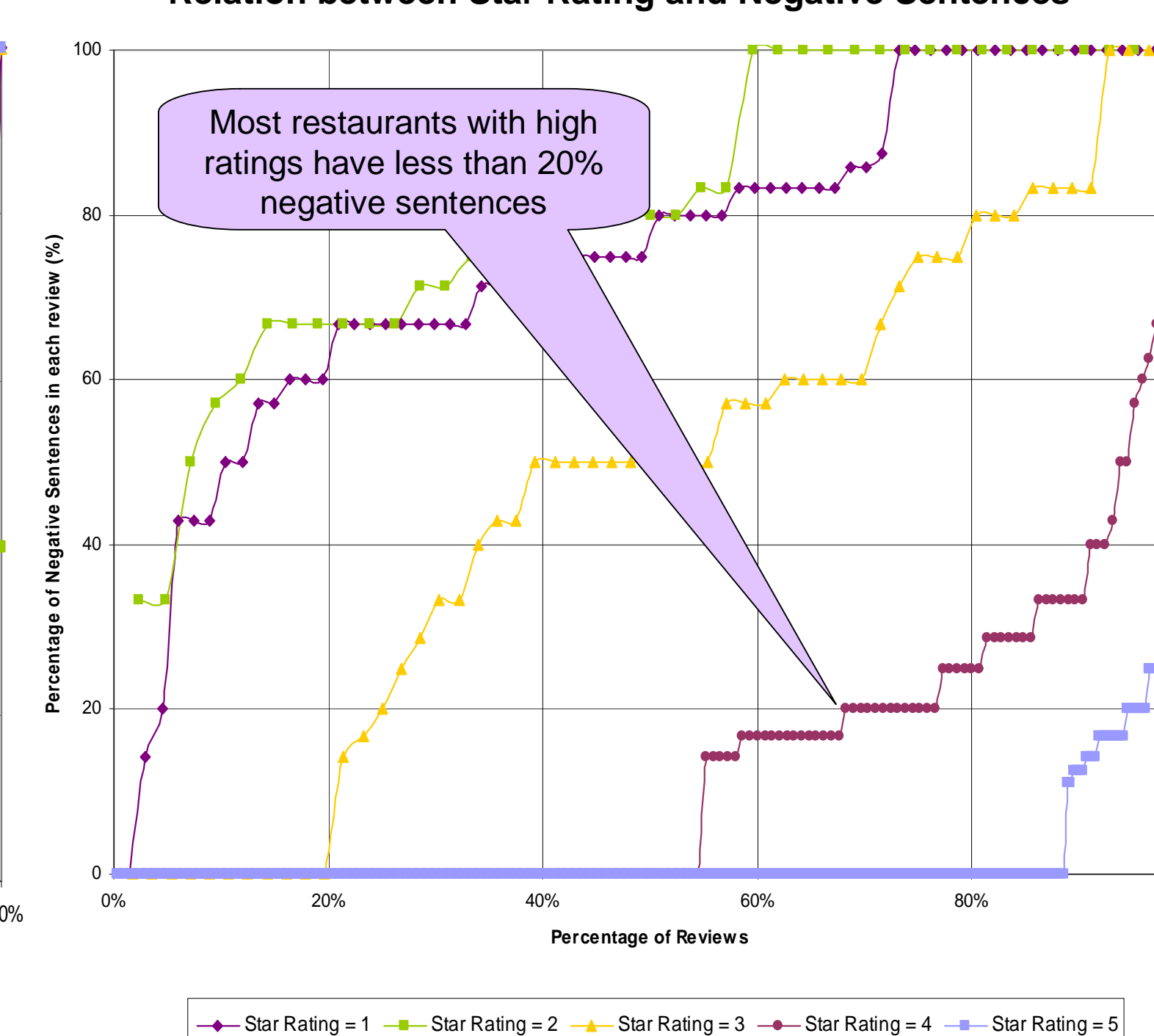


Correlation between structured metadata and textual reviews

Relation between Star Rating and Positive Sentences



Relation between Star Rating and Negative Sentences



Research Challenges

- **Structure Identification and Analysis**
 - Create data schema models.
 - Annotate training set and classify the dataset.
 - Analyze correlation between textual part of the review and the structured meta data.
- **Text and Structure Search**
 - Search reviews with emphasis on a particular aspect like food, price, ambience.
 - Allow complex searches and synonym searches like "Find all romantic restaurants".
 - Explore techniques that approximate both structure and content of query results.
- **Similarity Search in Social Networks**
 - Cluster users based on preferences of cuisine, price, restaurant location (meta data).
 - Cluster users based on the importance users attach with the different aspects of a restaurant like food quality, ambience, service (derived from textual data).
 - Perform restaurant based clustering.
 - Use collaborative filtering to make a recommendation system.

Examples and Challenges

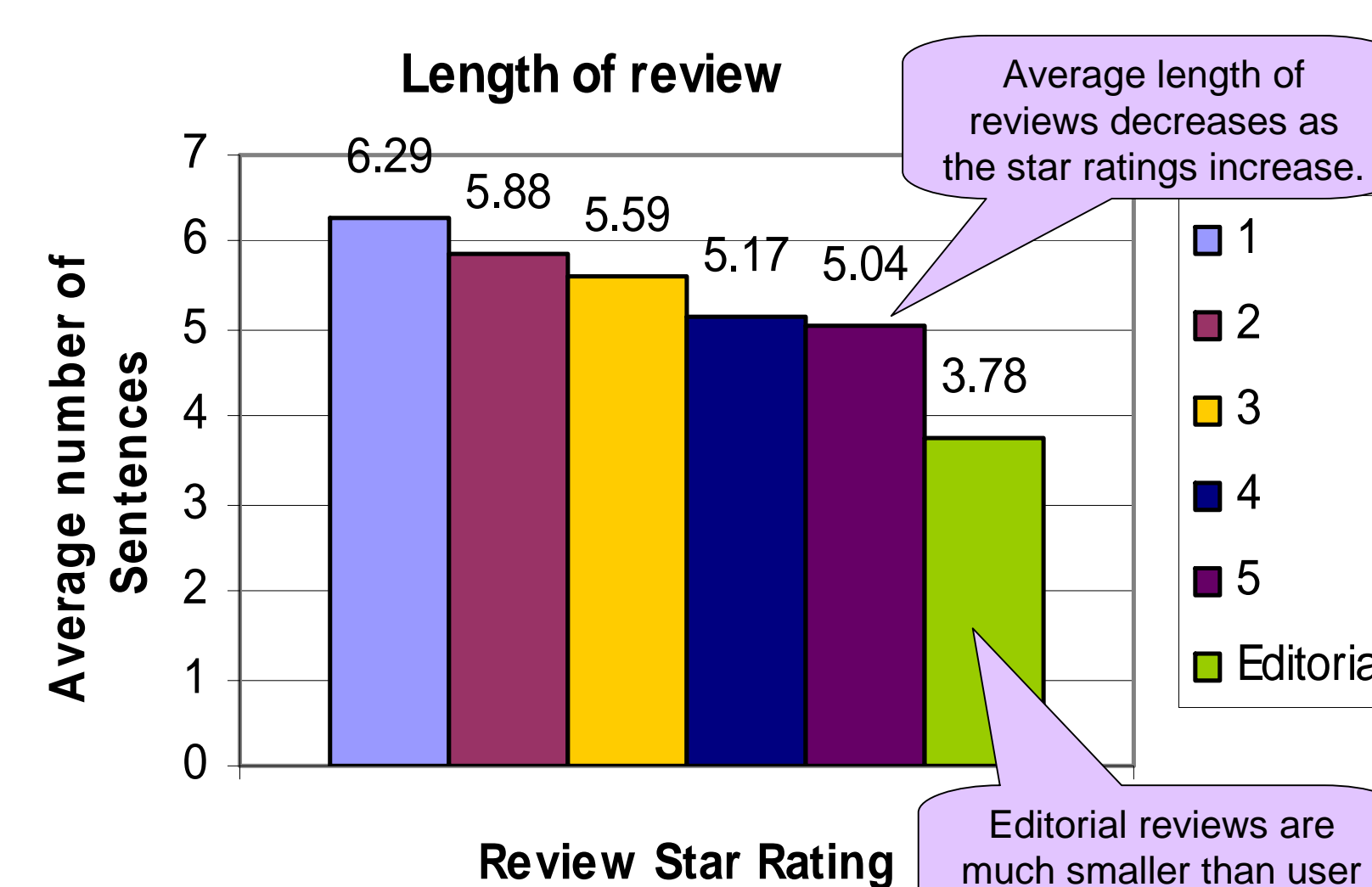
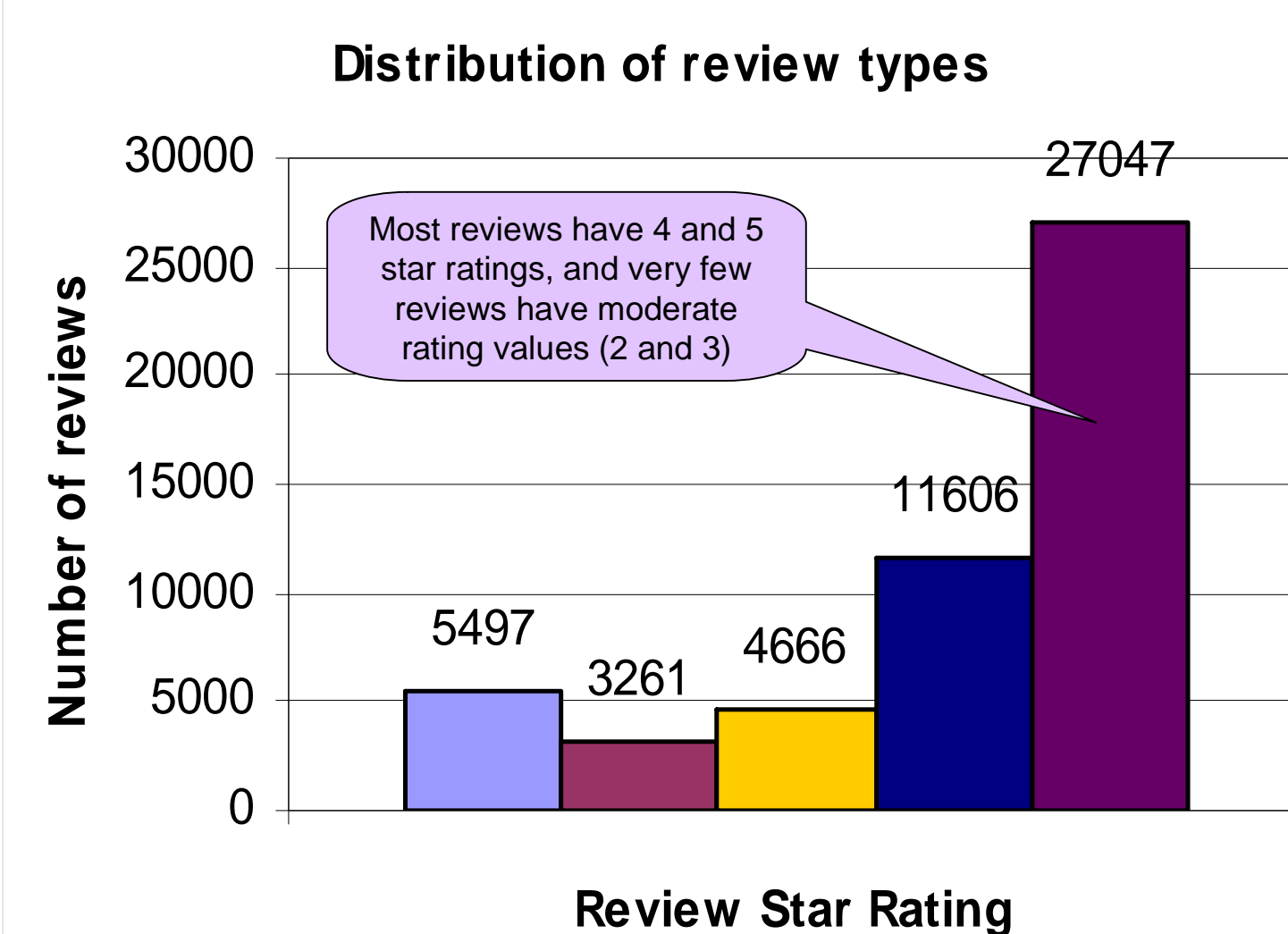
Restaurant Name : **Bandol**
"Tiny **dessert** was \$ 8.00 ...just plain overpriced for what it is."
"The mussels were fantastic and so was the **dessert** ...definitely going to be back very soon."
Keyword search is ineffective

Restaurant Name : **Pinky Pony**
"I had been searching really hard for a restaurant in New York where I could really **feel unwanted and ignored** and I finally found it! The **staff ignored** my friends and I the entire time we were there... **You guys are awesome!**"
Sarcasm makes it hard to determine the sentiment

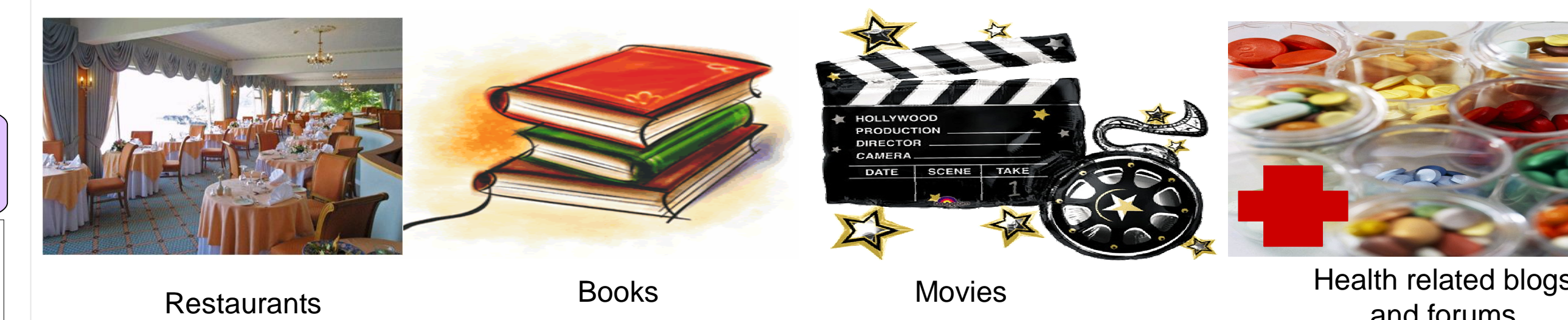
Name : **Heartland Brewery Midtown** (Rating 3)
"ALWAYS full with tourists ... But this place is a well-oiled machine so **they know what they're doing**. I've been here a bunch of times now and the **service is always outstanding**. The **food is above average**..."
Name : **Café Noir** (Rating 3)
"...**people who work there are evil and incompetent!!** The **service was terrible**, we had to wait for everything and ask several people for the same thing before we were allowed to be served... **The whole set up is truly unprofessional** ..."
Star rating is not a correct indicator of restaurant quality

Restaurant Name : **Café Noir**
"My friends and I showed up on a Sat night without a reservation, we had to wait at the bar for a little while, but the manager was so nice and made our wait a great experience. Bar was a little bit crowded, but **these five girls know how to have fun!!** it was a little hard to understand the waitress and she seemed to have little patience with our questions. We had to ask the manager and he was very helpful."
Anecdotal information does not help to evaluate the restaurant

Review-level analysis



Applications



Conclusions

The URSA project helps to classify, analyze, and search web user reviews by augmenting textual-based reviews with structural information. This information in conjunction with the meta data in reviews allows for better searching and collaborative filtering via exploring associations in a social network setting.