Using Syntagmatic Relations to Examine Semantic Communities in User Generated Book Reviews

Kyunghye Yoon, Adam Bezdicek
St. Catherine University
2004 Randolph Ave, St Paul, MN 55105
kyyoon@stkate.edu, ambezdicek@stkate.edu

ABSTRACT
While several studies have suggested that implicit semantic communities can be identified within user generated content, the majority have focused on identifying these communities via networks of “tags” within folksonomies. This method can be problematic due to the lack of linguistic context inherent in tags as collections of discrete terms. This study attempted a different approach to this problem by analyzing a corpus of user generated reviews for popular science books taken from Amazon. It searched the reviews for patterns of shared core semantic meanings as constituted by syntagmatic relations between “topic” and “comment.” Preliminary results showed that (1) core semantic meanings expressed via “topic” and “comment” relations were salient across reviews for a given book and across books within the genre and (2) specific patterns of core semantic meanings suggested the presence of different user communities constituted through similar manners of commenting on the books. The results provide insight into the social ways that readers accumulate shared meaning around books through repetition, interpretation, annotation.

Keywords
semantic communities, user generated book reviews, discourse analysis, syntagmatic relations.

INTRODUCTION
The rise of Web 2.0 has allowed for the emergence of new forms of community constituted through collaborative creation, curation, and sharing of information. While the boundaries of such communities are often explicitly defined via the shared usage of a particular website or technology, the formation of networked user groups dedicated to shared interests, and overt methods of policing community identity, the highly collaborative nature of Web 2.0 also allows for the formation of implicit communities defined by the shared adoption of similar practices, vocabulary, preferences, and behaviors (Vakali & Kafetsios, 2012; Cattuto, et al, 2008). The identification of implicit community features has become increasingly important as technologies that allow for user classification and descriptions of documents such as user tagging and user generated reviews gain in popularity. Analyzing the “semantic space” that users create around these documents, or the way that meanings assigned to a given document are created, negotiated, and shared, can provide insight into the dynamic emergence of communities constituted through networks of shared meaning (Cattuto, et al, 2008). Existing studies have primarily examined such “semantic communities” via analysis of networked links between topical tags in folksonomies (Li, et al, 2010; Xi, et al., 2014). This means that relationships between terms often have to be guessed at due to the tag’s lack of the linguistic context needed to fully understand its usage and meaning. Due to their greater capacity to convey contextual information in natural language, focusing on semantic communities within user generated book reviews might allow a better understanding of the relational meanings between terms that tags alone cannot capture.

While studies have examined the way that specific communities of user generated book reviewers implicitly enforce standards of practice through “helpful” and “unhelpful” ratings (Krestel & Dokoohaki, 2015; Ghose & Ipeirotis, 2011), few have focused on how the semantic content of these reviews might constitute shared ways of reading, making sense of, and evaluating books. This study attempted to address this gap, taking a discourse analysis approach to user generated online book reviews that views them as embodied “social actions” whose collective meanings cannot be easily reduced to discrete collections of words (Van Dijk, 1985; Talja, 1999). In doing so, it will identify core semantic meanings expressed via syntagmatic relationships, analyzing the patterns by which these meanings both remain consistent and change across reviews, books, and genres. Understanding these patterns of networked semantic relations can provide insight into the communities of practice that revolve around specific books and genres.

ASIST 2015, November 6-10, 2015, St. Louis, MO, USA.
SYNTAGMATIC RELATIONS IN DISCOURSE ANALYSIS

Discourse analysis is concerned with the way that meaning is informed and shaped by the context in which socio-linguistic interactions occur. It focuses on syntagmatic relations (i.e. words related only through context of use) rather than paradigmatic relations (i.e. words related via an abstract, codified, structure). Whereas paradigmatic relationships are defined a priori, as in the formal structure of an ontology, syntagmatic relationships can only be defined a posteriori, originating in the actual occurrence of terms within texts (Khoo & Na, 2006; Peters & Weller, 2008; Stock, 2010). Several studies have suggested that understanding how individuals employ syntagmatic relationships can illuminate how they classify information in online environments (Yoon 2012), create formalized structures of meaning through indexing (Olsen & Wolfram, 2008), and represent information at the conceptual level (Green, 1995b; 1995c; Khoo & Na, 2006). At a basic level, syntagmatic relationships create core semantic meanings within texts via relationships between “topic,” or that which the textual unit is about, and “comment,” or that upon which the text elaborates (Jakobson, 1963; Yoon, 2012).

Breaking down user claims into syntagmatic topic/comment relationships provides a systematic process for analyzing how the topic being addressed may be consistent across individual reviews, yet its meaning might change subtly or develop further depending on the comments made about it.

STUDY DESIGN

This study asked two questions:

R1: Is it possible to identify semantic cores of user meaning in terms of topic-comment relations that are salient across individual user book reviews in a way that better conveys the book’s content?

R2: If so, does such meaning convey the nature of specific book related communities that user generated reviews accommodate?

A pretest was conducted using reviews written for cook books and young adult novels, genres chosen for their relative simplicity. After the pre-test, a corpus of 200 reviews was purposively gathered from Amazon (amazon.com) by selecting the top 20 “most helpful” user reviews for 10 popular science books. The body of the reviews along with relevant metadata (i.e. usernames, star ratings, helpfulness ratings, and date written) were archived and coded in QDA Miner.

Review text was coded at the claim level, or a unit of text constituting a core semantic meaning comprised of a topic and a comment made about that topic (Mancini and Shum, 2006). This sometimes constituted a single sentence, but more often than not several claims (i.e. topic/comment couplings) could be derived from a given sentence. Coding and data analysis were done in three phases. In the first, coding categories were generated inductively and iteratively in order to code the broader topics being addressed in the claims. Categories were allowed to be as specific as was required to account for the full content of the review, but as coding progressed were refined into hierarchical groupings of categories and subcategories. For example, a claim describing the book’s thesis would be coded as “content-thesis,” with “content” referring to the broad category the claim addresses and “thesis” the sub-category.

The second phase of analysis was done primarily with the core meanings of the claims users made about the book’s content across reviews for a single book especially focusing on how reviewers conveyed the book’s thesis/theses within the context of their own evaluations and critiques. Claims about a book’s thesis were assessed in terms of topic/comment and compared across reviews to look for consistent patterns in meaning and expression. Claims were aggregated into similar groups when variation was minor enough across reviews for their core semantic meanings to remain consistent. This provided an overview of the full range of claims pertaining to a book’s thesis/theses conveyed in the reviews.

In the third phase of analysis, claims expressing similar core meanings were evaluated within the greater context of the reviews to see if reviewers who displayed similar expressions of the book’s thesis shared any further semantic features in common. Did reviews that shared thesis claims also display other features such as references to other reviews, specific types of personal statements, pervasive negative critiques, and etc., which might be indicative of implicit semantic communities? Are these community features salient across books within a single genre or are they only prevalent within a single book?

PRELIMINARY RESULTS AND DISCUSSION

Coding is still in progress, but initial analysis indicated that comments reviewers made (i.e., user meanings) could be grouped into 5 broad topical categories. These included comments about (1) the author, (2) the book as an object, (3) the book’s content, (4) the reviewer, and (5) other reviews. Within these 5 main categories multiple subcategories were identified, several of which were more prevalent in popular science reviews compared to pretest genres. Three of the most noteworthy genre specific subcategories included content-thesis statements (i.e. comments about the arguments that the book advances), reviewer-critique statements (i.e. comments constituting the reviewer’s positive or negative critique of the book’s contents), and reviewer-authority statements (i.e. comments about the reviewer’s expertise or familiarity with the book’s topic usually employed to bolster their critique).

While content-thesis statements often varied in terms of the specific phrasing and wording that users employed to recount the author’s arguments, in many cases core semantic meanings could be identified that were consistent across reviews within a single book. For example, 36 content-thesis statements were made by users in reviews written for Steven Levitt’s Freakonomics. Of these, the
most prevalent core semantic meaning, reflected in 9 of the 20 reviews, was users’ restatement of the book’s thesis that crime rates (topic) were affected by the passage of Roe v Wade (comment). Other statements of the book’s theses shared similar topic statements, but were too divergent in comments about those topics to be salient across reviews, which needs further in-depth analysis.

Similar patterns of core semantic meanings were identifiable among reviewer-critique statements (i.e. comments concerning the reviewer’s positive and negative critiques of the book’s content) and reviewer-authority statements (i.e. comments about the reviewer’s familiarity and expertise with the book’s topic). Content-thesis statements were often made only as part of the reviewer’s personal critique and juxtaposed with personal comments about the reviewer’s life, experience, and approach to reading. This highlights the fact that reviewer restatements of a book’s content (e.g. content-thesis) are always filtered through their own interpretations, annotations and personal perspectives. The reviewers’ intermingling of the explicitly personal (i.e. statements about their lives and approaches to reading) with reiteration, interpretation, and annotation of the book’s content illustrates how books can act as semantic spaces where their meanings are recreated and shared through personal and social actions. Further examination of these patterns in core semantic meanings could potentially be used to identify specific sub-communities of users who maybe be predisposed to speaking about, evaluating, or reading books differently.

The poster will outline some of the more salient patterns of semantic meaning that held across reviews and books in order to further discuss how these patterns can be used to understand different communities of readers and reviewers.

REFERENCES