

Self-Reflection, Perception, Cognitive Semantics: How Social is Social Tagging?

Abstract: In social tagging, cultural forces lead to conformance. In this study of self-reflective tags, using a sample from *Delicious.com*, cognitive semantic lenses are applied to reveal the fluidity of associative structure. The clustered tags represent classification, socially generated by bandwagon effect, with variation attributed to cognitive scanning, primarily conceptual blending.

Résumé : Dans le monde de l'étiquetage social, les forces culturelles entraînent la conformité. Dans cette étude sur les étiquettes autoréflexives basée sur un échantillon provenant de *delicious.com*, la fluidité des structures associatives est révélée sous l'œil de la sémantique cognitive. Le nuage de mots-clés représente une classification générée socialement sous l'impulsion du regard des autres, dont les variations s'expliquent par le balayage cognitif, principalement l'homogénéisation des concepts.

The phenomenon of social tagging has seen rapid growth as the technique has spread across the World Wide Web, especially finding acceptance in social Web applications, sometimes referred to as Web 2.0. Research into the actions of taggers has yielded a fair amount of knowledge about the tags assigned and some information about the behavior of taggers. An interesting observation is that cultural forces are at work alongside the taggers such that their behavior conforms to observable social patterns. In particular it is becoming apparent that much tagging is self-reflective. This abstract describes part of a study of self-reflective tagging in which cognitive semantic lenses are applied to enhance our understanding of the fluidity of associative folksonomic structure.

1. Background: Social social-tagging

Munk and Mørk (2007a and 2007b) discuss the manner in which users of social web services become invested in the system, and then participate in a bandwagon effect, by reusing already popular keywords as tags. This forms a “horizontal social categorization” (2007a, 16), in which users follow the bandwagon so their own tags will display alongside popular tags (p. 18). Consensus among taggers determines which few but popular specific tags will follow a power law. Most tags are unique but infrequently used; a few popular tags dominate the consensus and gain permanent visibility so long as the consensus is maintained. Cognition generates these dominant tags by intuitively stipulating denotative categories (2007a, 22).

Kipp and Campbell (2006) found that the most frequently used tags account for approximately 25% of all tags. They also discovered that distance between otherwise similar terms likely reflected differing user groups. Kipp (2008) discovered evidence of a dynamic relationship between taggers and their resources; contextual issues seemed to be reflected in the affective and subjective tags. Wolfram, Olson and Bloom (2009) found no differences in inter-indexer consistency density across disciplines, which suggests once again the highly personal aspects of tagging behavior.

Most tagging research has been focused on analysis of the tags, with an inferential understanding that those who contribute tags do so to provide shared indexing. Some

authors go so far as to posit a distinction between taggers who tag for themselves and those who tag for a shared environment. Ying Ding et al. (2009, 2391) make just such an inference when they discuss the dilatory effect of natural language on shared retrieval of tag data. However, it seems clear that some proportion of tagging is social in a different way—not egalitarian indexing for the masses, but rather more a form of personal assertion. Tags used in this manner are not intended primarily for indexing, or at least not for indexing for information retrieval. Rather they are left behind as markers to indicate ownership (or at least finder’s rights) of web resources. Such tagging is, of course, highly self-reflective.

1.1 Background: Cognitive semantics

Self-reflectivity is a hallmark of cognitive semantic activity, in which we see evidence of meaning interpreted through conceptualization. Langacker (2005) discusses the linguistic bases of cognitive semantics through aspects of a process he calls “scanning.” In scanning, the human engages in construal, through a series of mental processes that rapidly embraces essentially fictive statements by making sense of them collectively. Fictive scanning involves a sort of mental gymnastics that allows us to escape the bonds of temporality and space in order to colorfully portray experience in language. In such cases we might engage in any or all of several fictive portrayals. We might say “my teacher’s books keep getting longer” to indicate it seems the teacher is reading more all the time, when in essence what we have said is that the books are increasing in length. This activity is a commonplace feature of everyday language.

Wherever social tagging displays self-reflectivity we might expect to find fictive scanning among the tags assigned. Understanding this particular usage of tagging can help us understand the social implications of Web 2.0 conceptual spaces. The most promising research to point in this direction to date is that by Olson and Wolfram (2008) in which syntagmatic relationships—non-semantic, context-dependent, syntactic relationships are discovered as classes of similarity among tag-groups. Different taggers, encountering similar perceptual experience, produce syntactically similar (if semantically divergent) descriptors.

2. An Empirical Study

Thus we report a rigorous study of tags derived at random from *Delicious.com*, and analyzed using the cognitive semantic lenses developed by Langacker. The present study is the second phase of analysis of a sample drawn by Smiraglia (2010) used to discuss the presence of perceptual issues in the application of tags. Collected tags of individual taggers were drawn from the website *Delicious.com*.

2.1 Results

For the 42 sites in the sample, 11,378 tagging instances were recorded, with a mean of 271 taggings per site. 1730 individual tags were used (mean per site was 41). Each tag was used, on average, 5.8 times, although the range was quite wide, from 1– 118. These results are consistent with those reported in the prior tagging studies.

2.2 The Taggers

Emphasis on self-reflective tagging suggested analysis should begin with a discussion of the taggers observed in this study. There were 3582 taggers contributing to the 42 sites in this study. The mean number of taggers per site was 85.28, and the mean number of tags assigned per tagger was 3.36. The distribution of tagging follows a power law with 921 appearing more than once, and 2661 (74%) appearing only once. An even smaller

group—115 tagged more than twice, and only 18 tagged 5 or more times. To discover how truly “social” the tagging was, multi-dimensional scaling was used to analyze the incidence of co-tagging, which is the co-occurrence of two taggers on the same site, among these 18 most-prolific taggers. A two-dimensional map (Smiraglia 2010, 67) shows two clusters populated by web-designers, who assign mostly descriptive tags, and programmers, whose tagging is more self-reflective. Cultural “norms” set by consistent action within the two groups provides the backdrop for self-reflection. We observe a fair amount of social coherence (otherwise known as inter-tagger consistency) among them, which in turn suggests the presence of common cultural dynamics. Self-reflection takes place against the backdrop set by the culturally normative tags assigned by the relatively consistent prolific taggers. Less than 1% of the tags fall into the category of “fictive” cognitive semantics.

2.3 The Tags

We expected to find eccentric affective tags in small numbers throughout the sample. Ego-centricity means the occasional outburst might clearly stand out from the crowd. Indeed, very small numbers of tags in the 4% that were affective can be attributed to this type of behavior. However, a bandwagon effect was observed as taggers strained to join the most frequently used tag clusters. The majority of taggers assigned a few common terms, which dominated the tags for each site. Most of the remaining tags could be clustered either linguistically or semantically. Thus, the clustered tags constitute a classification—a *social* classification—of each site.

Specifically, the results are:

- Typically there are only a few clusters—the mean was 4.6.
- The two largest clusters contained nearly half (45%) of the total number of tags.
- All of the clustered tags contained, on the average, 63% of the total tags.
- The mean number of tags in a cluster was 42; 2/3 of the tags found their way into clusters.

The clustered tags represent classification, socially generated by the bandwagon effect, for each site. Much of the variation in each cluster can be attributed to “cognitive scanning,” in which people use shorthand, substitute verbiage, or other linguistic shortcuts to express themselves. Some of the variation is syntagmatic. Most of the observable variation falls into the category of conceptual blending, in which concepts belong to classes with fuzzy and overlapping boundaries.

3. Conclusion: Perception, Scanning, and Social Classification

The social action of tagging to get on the bandwagon creates a distinctly social classification for each website. Semantic divergence and syntagmatic conformance demonstrate cognitive semantic activity taking place between and among different user groups tagging the same sites. The taggers collectively are generating a classification with a social basis. There is still a large degree of variation in the singleton terms, but somewhere between half and two-thirds of the tags assigned can be attributed to one or more of the essential classes for each site. The clusters are not mutually exclusive, demonstrating that a natural classification is not necessarily either hierarchical, or mutually exclusive. But it does remain collectively, potentially, exhaustive. Warrant becomes a new issue in such a classification, because there is no accountable literary warrant—rather, warrant is cultural.

The collectivity of terms constitutes less a taxonomy than a social classification; the contributive work of the socially-engaged participants. It would be wise now to investigate tagger motivation. Cultural norms among the taggers in a fairly well-defined region such as *Delicious.com* fuel the classificatory activity; alliances among the taggers generate demonstrable social clusters that reflect the resulting divergence of terms within the socially-generated classificatory structures. How pervasive is this behavior among social taggers in more diverse environments?

4. References

Kipp, Margaret E.I. 2008. @toread and Cool: subjective, affective and associative factors in tagging. *Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British Columbia, Vancouver, June 5-7, 2008*. http://www.cais-acsi.ca/proceedings/2008/kipp_2008.pdf.

Kipp, Margaret E.I. and Campbell, D. Grant. 2006. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. *Proceedings of the 2006 Annual Meeting of the American Society for Information Science and Technology, Austin, November 3-8, 2006*. eprints.rclis.org/archive/00008315/.

Langacker, Ronald W. 2005. Dynamicity, fictivity, and scanning: the imaginative basis of logic and linguistic meaning. In Pecher, Diane and Rolf A. Zwaan eds., *Grounding cognition : the role of perception and action in memory, language, and thinking*. Cambridge : Cambridge Univ. Pr., pp. 164-97.

Munk, Timme Bisgaard and Mørk, Kristian. 2007a. Folksonomy, the power law & the significance of the least effort. *Knowledge organization* 34: 16-33.

_____. 2007b. Folksonomies, tagging communities, and tagging strategies-an empirical study. *Knowledge organization*. 34: 3.

Olson, Hope A. and Wolfram, Dietmar. 2008. Syntagmatic relationships and indexing consistency on a larger scale. *Journal of documentation* 64: 602-15.

Smiraglia, Richard P. 2010. Perception, knowledge organization, and noetic affective social tagging. In Gnoli, Claudio and Mazzocchi, Fulvio eds., *Paradigms and conceptual systems in knowledge organization: proceedings of the Eleventh International ISKO Conference 23-26 February 2010, Rome, Italy*. Advances in knowledge organization 12. Würzburg: Ergon Verlag, pp. 64-70.

Wolfram, Dietmar, Olson, Hope A., and Bloom, Raina. 2009. Measuring consistency for multiple taggers using vector space modeling. *Journal of the American Society for Information Science and Technology* 60: 1995-2003.

Ying Ding et al. 2009. Perspectives on social tagging. *Journal of the American Society for Information Science and Technology* 60: 2388-2401.