Membership kind and format modality in schema.org categorization

Francesca Zarl, Walter Gerbino

Psychology Unit “Gaetano Kanizsa”, Department of Life Sciences, University of Trieste, Italy

Keywords: categorization, classes, membership, ontology, schema.org.

Ontologies in computer science are conceptual models that are augmented by formal axioms [1]. Schema.org is an example of a web ontology that aims to represent goods online, by defining a shared specification of conceptual elements for data interoperability [2]. Historically, the design of such ontologies has been predominantly governed by membership-based rules. Recently, the consideration of the human cognitive functioning is becoming increasingly significant in ontologies concepts representation [3]. As emerged from the literature analysis, typicality [4-7] and format [8,9] should be considered in ontology concepts representation.

Based on that, the aim of this study is to evaluate participants’ responses regarding the concepts representation in schema.org, considering the membership kind (central vs borderline) and format modality (text vs image).

We ran a computer-based experiment in which participants had to categorize random instances (represented using either text or image formats), by answering Yes or No to an ISA question of the form “IS exemplar x An instance of category c?”. After the Y/N answer the participant evaluated his/her confidence on a 1-4 scale (with 1= absolutely certain and 4= absolutely uncertain).

Eleven categories were randomly selected from schema.org. For each category, a balanced set of questions was generated according to a Membership (2) × Format (2) within-subject design. Exemplars might be either central or borderline (the two levels of Membership), and either an image or text (the two levels of Format). The aim was to measure the influence of these variables on categorization.

First, we ran separate one-way analyses of responses to ISA questions using Membership and Format as independent variables. Y/N and 1-4 confidence rating data were analyzed according to the unequal variance SDT model. A significant effect of Membership emerged, confirming that ISA judgments were more accurate for central than borderline instances; whereas there was no effect of Format. The response bias was statistically significant, though small, in both analyses: participants showed a liberal bias in the Membership analysis (higher yes rate for borderline than central exemplars) and a smaller bias in the Format analysis (higher yes rate for image than text).

Since the response bias was small, a Membership (2) × Format (2) repeated-measures ANOVA was carried out on percent correct. We obtained a significant interaction: answering ISA questions involving borderline instances were equally difficult for image and text instances (no effect of Format). On the other hand, questions involving central instances was easier for images than text.


