Extracting Gender Stereotypes with Latent Semantic Analysis
Marie Gustafsson¹, Sverker Sikström²
¹Department of Psychology, Stockholm University, Sweden, ²Lund University Cognitive Science, Sweden

Abstract:
New computational method to study value and stereotypes

Current research on female and male stereotypes has problems in explaining how stereotypes arise, sustain and are shared between people. Empirical studies have found that stereotypes are reproduced and enhanced in communication (Lyon & Kashima, 2003) and media articles could therefore be interesting to examine. Latent semantics analysis makes it possible to analyze text as an empirical record of human behavior. The analyses are completely data driven and not influenced by prior values. The technique makes it possible to analyze larger text volumes than humans can read.

The findings show that females are higher evaluated than men, and that females are more homogeneously described.

Introduction:
Several theories and empirical studies have tried to explain why there are so few women in top leadership positions. Many theories focus on different gender stereotypes and different evaluations of women and men. The male stereotype is often described as agentic (being more aggressive, dominant and having leadership abilities) and the female stereotype is described as communal (being more listening, relation oriented and sensitive). As a consequence females will not get leadership positions because they will be judged as having lower leadership abilities than men and females could also be negatively evaluated if they break their gender role (Eagly & Karau, 2003). Although, women in general are often positively evaluated; female leaders could be stereotyped as competent but cold (Fiske et al., 2003). In this study we analyze gender stereotypes using media articles as an empirical record of human behavior to examine the female and male stereotypes in Sweden.

Method:
Latent semantic analysis (LSA) analyzes meaning by co-occurrences. Every word is analyzed through the context where it is represented in the text. The data consist of 100 000 articles published in Swedish business media during year 2000-2006. All articles include the word “she” or “he”, together with “business leader”.

Steps:
1. Transform the text corpus into a distributed dimensional semantic representation using LSA and the SVD algorithm.
2. Locate valence by multiple linear regression from semantic representation to 288 sex-neutral words ranked for valence.
3. Predict non-ranked sex specific words for valence.
4. Cluster documents into topics.
5. Find gender stereotypical words by listing words most closely associated to “she” and “he” within each topic.

Results:
Females are generally higher valued:
Mean valence Female (she): +0.765 (±1.27) N 21.442
Mean valence Male (he): +0.692 (±1.32) N 103.254

The overall higher valence for females is due to the fact that a majority of the “she”-articles are represented in high valence topics (62% of the “she”-articles).

The female topics are similar to the female stereotype in their personal and life oriented content rather than business oriented content.

Female words are more often presented with first names and males are more often presented with their last names. Gender marked words are more common to describe females (i.e. “the first female CEO”). The content analysis (fig. 3) of the associated words show that there are significant differences in how females and males are presented, p<0.001.

Discussion:
Outgroup homogeneity effects have previously been found in studies that compare descriptions of ingroups compared to outgroups. Business media could be seen as a male dominated area where masculinity is the norm. The group should therefore be “he” and a possible explanation of the homogeneity effect of the words associated to the outgroup “she”.

The higher frequency of first names associated to “she” could be interpreted as females lack status. Previous studies have found that last names is a marker for status and hierarchy and that first names are used in more personal communication and to address lower status.

The higher evaluation of females in this corpus contradicts previous studies on female leadership. The higher valence for females could be interpreted as a result of the “women-are-wonderful-effect” but it could also be an effect of the focus on equality between the sexes or about being politically correct. There might also be cultural differences between the US and Sweden in how gender and leadership is evaluated.

Latent Semantic Analysis is a promising technique for evaluation of shared values and attitudes. Future research may consider other dimensions to understand gender and leadership stereotypes.

References