"She" and "He" in news media messages:

Pronoun use reflects gender stereotypes

Marie Gustafsson Sendén

News media aim at unbiased presentation of social groups in media reports. For example, in the Handbook of Journalism, published by Reuters it is stated:

> … A person's gender, race, religion, nationality, sexual orientation or marital status should not be cited unless it is relevant to the story. Even then, consideration must be given to where in the story such information needs to be placed. It is wrong to assume that police, firefighters or soldiers are men…
> (Retrieved from http://handbook.reuters.com/?title= G#gender)

Although the directives in the handbook are clear, previous research has shown that news media in general tend to be biased when reporting on different social categories, such as gender or minority groups (Dixon & Linz, 2000; Geschke et al., 2010; Koivula, 1999; Rudy et al., 2011). In comparison to advertising and fiction, where blatant stereotypes are common (see for example; Lauzen et al., 2000; Mager & Helgenson, 2011), news media may include more subtle ways to represent women and men that confirm prevalent gender beliefs. Since news media explicitly aim at unbiased reports, such subtle biases may be more difficult to detect, and their influence on people's notions of women and men may therefore go largely unnoticed.

In comparison to research on advertising and the entertainment media, analyses of gender biases in news media are scarce. In a review in 2010 of scientific articles focusing on content analyses of gender in the media, 93 studies investigated entertainment and advertising media, whereas four focused on media news (time span 1978-2009; Rudy et al., 2010). The few studies examining gender biases in news included comparisons of the ratio of women and men as news targets, qualitative analyses on how men and

women are presented, whether women and men occur in different news topics, and whether they receive the same or different types of questions (see for example; Koivula, 1999; Matud et al., 2011; Schwartz, 2011). In the current study, the question of gender biases in news media is approached by focusing on personal pronouns denoting women and men (*she* and *he*). Specifically, it is examined how gender pronouns in news vary with respect to frequency, and how they become associated with meaning. To this aim, we use news published in English from Reuters, which is the largest international news provider in the world (http://reuters.com). Reuters provides short media news articles which are reprinted and redistributed in magazines and internet services all over the world, which means they have a critical impact on people's worldview.

Pronouns are especially well-suited for research on language biases since they reflect important social categories (Gustafsson Sendén et al., in press; 2014) and are used with less cognitive monitoring than words denoting explicit categories (Pennebaker et al., 2003). This means that pronouns are used more automatically and unconsciously than explicitly defining categories, such as men or women.

The analyses are conducted by latent semantic analysis (LSA; Landauer, 1999; Landauer et al., 2007), which is a multi-dimensional computerized method that analyses meaning by examining how words are used in specific contexts. The LSA is completely data-driven and allows analysis of large amounts of data where manual coding would be highly time consuming. In this study, 400 000 articles are included in the analyses. Using manual coders in such a project would not only take a very long time but also be extremely expensive. News reports may also contain subtle differences that would be difficult to detect by human coders. Previous studies have used corpus-based computerized analyses for analyses of frequencies of pronouns (see for example Baker, 2010, Twenge et al 2012). However, to understand the meaning of a word, qualitative methods are added (Baker, 2010). In comparison, LSA derives the meaning of a word by computational algorithms. Because the method is still rather unknown within psycholinguistics and gender research, I explain it further in the method section.

Gender Stereotypes

Gender stereotypes are shared beliefs about women and men, including both descriptive and evaluative aspects (Glick & Fiske, 1996; Fiske & Taylor, 1991) that are transferred in communication. The typical male stereotype

include traits such as *competent, active, assertive, confident, aggressive, ambitious and dominant*, whereas the female stereotypes include traits and behaviors such as being *warm, passive, empathic, affective, helpful, and kind* (Bakan, 1966; Cuddy et al., 2004; Eagly, 1987; Fiske et al., 2002). According to social role theory (Eagly & Steffen, 1984, 1986), gender stereotypes result from the distribution of women and men in social roles, for example, occupying different professions. When people see groups of people performing particular tasks, personalities and competences are attributed from the behavior to the social category. For example, if people see women (but not men) taking care of children, nurturance and warmth will be associated with women (Eagly & Steffen, 1984). In a similar vein, the media may function as an active agent in transmission of gender stereotypes (Bussey & Bandura, 1999), if presenting women and men in different roles, topics and in different numbers.

Gender biases in language

In the media, category labels are more common for women than men, and essentialising language might be reflected in the content of interviews with women and men (Rudy et al., 2010). Supporting this notion, research has found that interviews with male and female business leaders differ, such that interviews with male leaders focus on the topic (e.g., business), whereas interviews with female leaders focus on family relations, and being a woman (Lämsä & Tiensuu, 2002). A similar effect has been found in sports news (Koivula, 1999). Taken together, this also means that business-and sports-women may be more similarly depicted than business- and sports-men.

Another linguistic device that may convey attitudes is word categories, such as verbs, adjectives, or nouns (Semin & Fiedler, 1988, Carnaghi et al., 2008). For example, nouns have stronger inductive potential, and more strongly indicate group membership than for example adjectives do (Carnaghi et al., 2008; Graf et al., 2012). For example, describing somebody as "*NN is a homosexual*" is associated with stronger stereotype congruent inferences than "*NN is homosexual*". *A homosexual* inhibited also imaginations of stereotype incongruent behaviors or interest (Carnaghi et al., 2008). Verbs on the other hand are less associated with personal inferences (Semin & Fiedler, 1988) and instead focus on the situations and the behavior as such.

Using LSA, so-called semantic associates may be used to assess the associative meaning of *she* and *he* in media news. The semantic associates represent words that are either used in the same contexts as *she* and *he*, or occur

in similar contexts as *she* and *he*. In accordance with previous research, the second hypothesis is that the semantic associates of *she* will include more essentialising language as expressed by gender labels, nouns and adjectives than the associates of *he.*

Method

In studies of artificial intelligence there has been development of mathematical ways to analyze similarities between words and contexts that do not depend on human judges (Campbell & Pennebaker, 2003; Landauer et al., 2007; Pitts & Nussbaum, 2006). One such promising technique is LSA (Landauer & Dumais, 1997, Landauer 1998). LSA has been suggested as a theory and a method for how semantic meaning can be derived from the contexts in which words are used (Landauer & Dumais, 1998; Landauer, 1999; Landauer et al., 2007). The exact calculations in LSA are beyond the scope of this paper, but suffice to say, LSA relies on singular value decomposition, which is akin to factor analysis and multidimensional scaling (Landauer & Dumais, 1997; Landauer et al., 2007). Through complex computations, LSA learns about the meaning of a word from every encounter with it and from the composition of all the passages in which a word does not occur. "…the similarity estimates derived by LSA are not simple contiguity frequencies, co-occurrence counts, or correlation in usage, but depend on powerful mathematical analysis that is capable of correctly inferring much deeper relations (thus the phrase *latent semantic)"* (Landauer et al., 1998, p. 261). The computations in LSA suggest that the meaning of *she* and *he* can be imagined as the average meaning of all the passages in which *she* and *he* occur in the same corpus, and also from all the passages where these words do not occur (Landauer et al., 2007; Lenton et al, 2009). For detailed descriptions, I refer to methodological articles on LSA (see for example, Foltz et al, 1998; Landauer, 1998; Landauer & Dumais, 1997; Landauer et al., 2007), and articles where LSA has been implemented in analyses of expressive writing (Campbell & Pennebaker, 2003), or gender stereotypes (Lenton et al., 2009).

Lenton, Sedikes and Brewer (2009), used a corpus based on fictional and educational literature to assess the distance in a semantic space between gender stereotypical roles and traits and gender labels. They found that gender stereotypical roles but not traits were associated with the matching gender labels, as indicated by the cosine between wordlists with stereotypical roles, traits and male and female gender labels. In the current study, the

focus is broadened from roles and traits to analyses of the closest semantic contexts in which *she* and *he* appear.

Material

To create a semantic space with media news, a corpus with approximately 800,000 thousand (404 megabite) Reuters news messages published in 1996 and 1997 was used. This dataset was released publicly by Reuters and it was one of the richest samples available for research when the current project was initiated (David et al, 2004; Rose, Stevenson and Whitehead, 2002). Reuters articles are short news messages covering subjects such as business, politics, entertainment and sports (www.reuters.com). The INFOMAP software (http://www.infomap-nlp.sourceforge.net) was used to implement the LSA algorithm. The quality of the semantic space is improved if the words used to build the matrix contain meaningful semantic content (Landauer et al., 2007), therefore, about 750 high frequency, non-content words (e.g., and, but, who) were removed. From the original dataset the 15,000 most common words were used to build a space with 100 dimensions. Two sentences (retrieved from www.reuters.com) are presented here to show how gendered pronouns might be used in a news context and the type of words included in the construction of a semantic space (words in bold).

"… NN can win her fourth title of the week in the 800 meters freestyle after golds in the 4x200 freestyle relay and the 400 and 1,500 freestyle, when **she** smashed the world record. If she wins as expected, NN will be the second woman to claim the 400, 800 and 1,500 golds in the same championships …";

" … faster by winning the 100m and 200m in world records that still stand. Two years later in South Korea it all went horribly wrong when **he** false-started and was disqualified from the 100, opening the door for NN. He roared back to take the 200 and relay golds, proving he is fully fit"

Semantic associates as defining the meaning

In comparison to other computational methods, LSA represents each word on a word-vector in a semantic space. This word-vector consists of the target word and weights for the 100 dimensions in the space, locating the word

in the space in comparison to all other locations of words. Calculating the distance (the cosine) between two word-vectors indicates how closely located these two words are in the space. Words that are closely located to each other are labeled 'semantic associates', and as such they should represent the meaning of the target word. For example, the semantic associates to *bad* are: *problem*, *performing*, *worse*; whereas the semantic associates to *good* are: *pretty, excellent* and *positive,* in the Reuters space. It should also be noted that these words can occur in the closest contexts of *bad* and *good*, but they also occur in contexts that are similar to each other. For example, words such as *doctor* and *physician* are not expected to occur in the same contexts or documents. However, these words become associates because they words are used in the same contexts (Landauer et al., 2007). In the current study, the semantic associates of *she* and *he* are analyzed to understand how *she* and *he* are used in the Reuters news contexts.

To sum up the hypotheses

1. *He* will occur more frequently than *she*.
2. *She* will be associated with more essentialising language than *he* as reflected by the semantic associates of *she* and *he*.

Results

The frequency analysis concerned the relative ratios of *he* and *she*. As predicted in the first hypothesis, *he* ($N = 81630$) outnumbered *she* ($N = 9075$) by a relative ratio of approximately nine instances of *he* for each instance of *she*.

The analysis of descriptive biases including essentialising language was assessed by qualitative interpretation of the semantic associates. Table 1 includes the semantic associates for the *she* and *he* contexts. The cosine is the angle between two word-vectors and defines how close each semantic associate is to *she* and *he*. The cosine can vary from -1 to +1, where +1 indicates that two words are used in exactly the same contexts. As can be seen, the semantic associates of *she* are gendered pronouns (her, herself), gender labels (woman, girl) and gender labels associated with family relations (mother, husband). The semantic associates of *he* include only one gendered pronoun (his), and no gender labels. In comparing word classes, there are four nouns among the semantic associates of *she* and zero nouns for *he*, whereas there are six verbs among the semantic associates of *he* and three verbs for *she*. Another difference that can be noted is that the semantic associates of *he* associates are related to action (made, make, adding, give),

whereas the semantic associates of She are more passive (admit, answer, refuse). The semantic associates also have higher cosines, which means that the semantic associates of *she* are more similar to the word *she* than the semantic associates of *he* are to the word *he.*

(1) Table 1. Semantic associates of 30 word contexts around *she* and *he.*

She-context	Cos	He-contexts	Cos
She	.66	he	.55
Her	.62	made	.52
woman	.57	clear	.51
mother	.56	make	.50
husband	.56	adding	.50
herself	.54	quickly	.49
Girl	.51	his	.49
answer	.50	who	.49
admitted	.50	giving	.48
refused	.50	give	.46

Discussion

In this study, gender pronouns were used to analyze the frequency of women and men in news media messages in Reuters news, and also how the meaning of *she* and *he* should be understood in these news messages.

The analyses comparing the relative frequencies of *she* and *he* showed that there was substantial male dominance. For each instance of *she* there were nine instances of *he.* Male dominance of a similar extent has previously been demonstrated in sports news (i.e., Koivula, 1999; Matud et al., 2011), but general news coverage typically shows two to four men for each women mentioned in the media (see for example; Armstrong, 2004; Schwartz, 2011). As we did not control for whether the occurrences referred to the same person or not, the large difference in ratios could either mirror the actual numbers of men and women in the news, or may indicate that men were given more space, for example by appearing in longer articles with more citations. In support of this notion, a Spanish study showed that news articles about women were shorter than articles about men, and that news about men was located in more salient positions on the newspaper page (Armstrong, 2004; Matud et al., 2011). If women are given less space

and/or fewer citations, it could be argued that the media consider women to be of less interest than men.

Hypothesis 2 included a qualitative analysis of the descriptive content of the words associated with *she* and *he*. The semantic associates extracted by the LSA represent the meaning of a word as it is used throughout the corpora of media news (Foltz et al., 1998; Landauer et al., 2007). The semantic associates of *she* were more gendered (e.g., mother, women and girl) than were the semantic associates of *he*, which supports the notion that underrepresented categories are more often labeled in accordance with their category (Kahneman & Miller, 1986; Kanter, 1977; Stahlberg et al., 2007). An unexpected finding from the qualitative analyses of the semantic associates was that the associates also reflected gender stereotypes of women as being passive and men as being active. For example, *she* associates included words such as *answer* and *admit*, whereas *he* associates included words such as *made* and *give*. This is in line with early theories on stereotypes representing women as passive and men as active (Bakan, 1966), and also later corpus based studies that found such differences when comparing adjective that are used in relation to man or women (Baker, 2010). However, we did not expect such a pattern to occur with only nine associates to the *she* and *he* contexts.

Taken together, these results further support the interpretation that news media frame their descriptions of women and men in accordance with gender stereotypes. One should remember that Reuters news is dominated by political and business news, and the associations of family relations to *she* in these types of context needs to be addressed more thoroughly in future research.

When the current project was initiated, this Reuters sample from 1996-1997 was one of the richest samples available for research. The imbalance in the ratio of women and men should have improved in media news today, since female representation in top positions and as journalists has increased. However, there is probably a long way to go until the numbers of *she* and *he* are equal. Men are still in the majority in top positions all over the world, and prejudice and bias mean that men are more often seen as experts than women (Desmond & Danilewicz, 2010, Rudy et al., 2010). Also in literature, which has more female than male readers (Tepper, 2000), there are still two instances of *he* for each instance of *she* (Twenge et al., 2012).

The findings imply that media both depict facts and stereotypes about gender. For example, it is true that women are scarce in top business or political positions, and as a consequence they should be less frequently depicted in those areas. However, the essential language in the semantic asso-

ciates of *she* and *he* indicates that the media are also responsible for contributing to stereotypes. Since the differences found here are repeated on a daily basis, and are spread in various forms of news media, the gender representations in news media are likely to maintain and reinforce gender inequity.

References

Armstrong, C. L. (2004). The influence of reporter gender on source selection in newspaper stories. *Journalism & Mass Communication Quarterly*, *81*, 139-154.

Armstrong, C. L. (2006). Story genre influences whether women are sources. *Newspaper Research Journal, 27,* 66-81.

Bakan, D. (1966). The *duality of human existence: An essay on psychology and religion*. Oxford, England. Rand McNally.

Baker, P. (2010). Will Ms ever be as frequent as Mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language, 4,* 125-149

Bragg, M. (2003). *The adventure of English: The biography of a language.* London: Hodder and Stoughton

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology*, 42, 210-218.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106,* 676-713.

Campbell, R. S., & Pennebaker, J. W. (2003). The secret life of pronouns: flexibility in writing style and physical health. *Psychological Science*, *14*, 60-65.

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. *Journal of Personality and Social Psychology, 93*, 1028-1039.

Collins, R. L. (2011). Content analysis of gender roles in media: Where are we now and where should we go? *Sex Roles, 64,* 290-298.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues, 60*, 701-718.

David, D. L., Yiming, Y., Tony, G. R., & Fan, L. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5, 361-397.

Desmond, R., & Danilewicz, A. (2010). Women are on, but not in, the news:

Gender roles in local television news. *Sex Roles, 62*, 822-829.

Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication, 50,* 131-154.

Eagly, A. H. (1987). Sex differences in social behavior: A social-role interpretation. Hillsdale, NJ: Erlbaum.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. J*ournal of Personality and Social Psychology, 46,* 735-754.

Eagly, A. H., & Steffen, V. J. (1986). Gender stereotypes, occupational roles, and beliefs about part-time employees. *Psychology of Women Quarterly, 10*, 252-262.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Fiske, S. T., & Taylor, S. E. (1991) *Social Cognition*. New York [etc.]: McGraw-Hill.

Foltz, W., Kintsch, W., & Landauer, K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25,* 285-307.

Geschke, D., Sassenberg, K., Ruhrmann, G., & Sommer, D. (2010). Effects of linguistic abstractness in the mass media: How newspaper articles shape readers' attitudes toward migrants. *Journal of Media Psychology: Theories, Methods, and Applications, 22,* 99-104.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491-512.

Graf, S., Bilewicz, M., Finell, E., & Geschke, D. (2012). Nouns cut slices: Effects of linguistic forms on intergroup bias. *Journal of Language and Social Psychology, 32,* 62-83.

Gustafsson Sendén, M., Lindholm, T. & Sikström (2014). Selection bias as reflected by choice of words: The evaluations of "I" and "We" differ between communication contexts but "They" are always worse. *Journal of Language and Social Psychology*, *33*, 49-67

Gustafsson Sendén, M., Lindholm, T. & Sikström (in press). Biases in news media as reflected by personal pronouns in evaluative contexts. *Social Psychology*.

Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences

in APA journals: 1965-2004. *Review of General Psychology, 10*, 377-389.

Kahneman, D., & Miller, D. T. (1986). Norm theory - Comparing reality to its alternatives. *Psychological Review, 93*, 136-153.

Kanter, R. M. (1977). *Men and Women of the Corporation*. New York: Basic Books.

Koivula, N. (1999). Gender stereotyping in televised media sport coverage. *Sex Roles*, 41, 589-604.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science, 7*, 161-164.

Landauer, T. K. (1999). Latent semantic analysis: A theory of the psychology of language and mind. *Discourse Processes, 27,* 303-310.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, New Jersey: Lawrence Erlbaum.

Lauzen, M. M., Dozier, D. M., & Horan, N. (2008). Constructing gender stereotypes through social roles in prime-time television. *Journal of Broadcasting & Electronic Media, 52,* 200-214.

Lenton, A., Sedikides, C., & Bruder, M. (2009). A latent semantic analysis of gender stereotype-consistency and narrowness in American English. *Sex Roles*, *60*, 269-278.

Lämsä, A., & Tiensuu, T. (2002). Representations of the woman leader in Finnish business media articles. Business Ethics: A European Review, 11, 363-374.

Mager, J., & Helgeson, J. G. (2011). Fifty years of advertising images: Some changing perspectives on role portrayals along with enduring consistencies. *Sex Roles, 64,* 238-252.

Matud, M. P., Rodríguez, C., & Espinosa, I. (2011). Gender in Spanish daily newspapers. *Sex Roles, 64,* 253-264.

Miller, D. T., Taylor, B., & Buck, M. L. (1991). Gender gaps: Who needs to be explained? *Journal of Personality and Social Psychology, 61*, 5-12.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547-577.

Rose, T., Stevenson, M., & Whitehead, M. (2002). *The Reuters Corpus Volume 1 – from yesterday's news to tomorrow's language resources.* Paper presented

at the Proceedings of the 3rd International Conference on Language Re-sources and Evaluation.

Rudy, R. M., Popova, L., & Linz, D. G. (2010). The context of current content analysis of gender roles: An introduction to a special issue. *Sex Roles, 62*, 705-720.

Rudy, R. M., Popova, L., & Linz, D. G. (2011). Contributions to the content analysis of gender roles: An introduction to a special issue. *Sex Roles, 64*, 151-159.

Schwartz, J. (2011). Whose voices are heard? Gender, sexual orientation, and newspaper sources. *Sex Roles, 64,* 265-275.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the Sexes in Language. In K. Fiedler (Ed.), *Social Communication.* (pp. 163-187). New York. Psychology Press.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Male and female pronoun use in US books reflects women's status, 1900–2008. *Sex Roles, 67*, 488-493.