A Gender Based Study of Tagging Behavior in Twitter

Evandro Cunha [1], Gabriel Magno [1], Virgilio Almeida [1], Marcos André Gonçalves [1], Fabrício Benevenuto [2]

[1] Computer Science Department, Federal University of Minas Gerais (UFMG), Brazil
[2] Computer Science Department, Federal University of Ouro Preto (UFOP), Brazil

MOTIVATION AND GOAL

• According to several linguistic studies, gender plays a key role in the process of language variation. **Men and women use language in different ways, according to the expected behavior patterns associated with their status in the communities.**

• We present a first description of gender distinctions in the usage of Twitter #hashtags and **conclude that gender can be considered a social factor that influences the user's choice of particular hashtags about a given topic.**

• This study aims to increase knowledge about **human behavior in free tagging environments** and may be useful to the **development of tag recommendation systems based on users' collective preferences.**
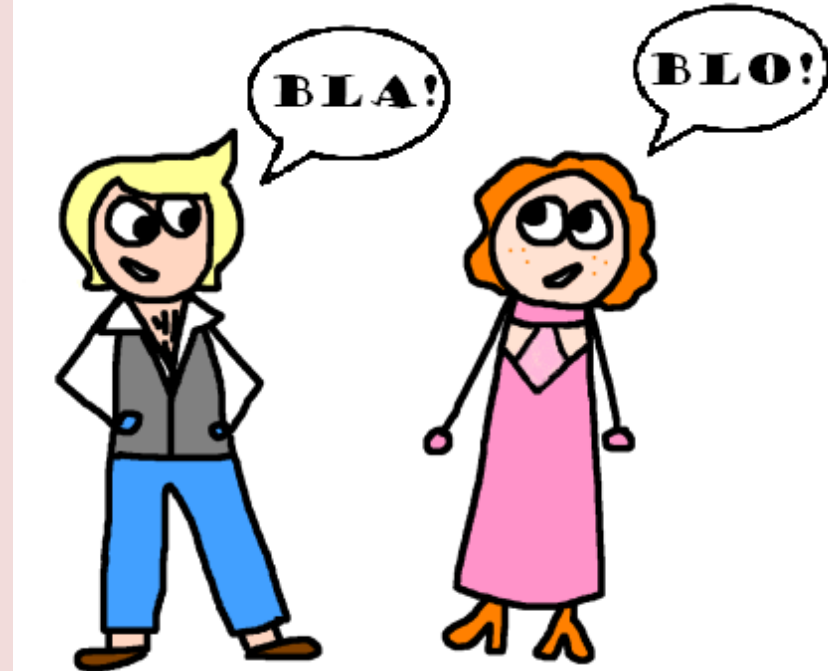
METHODOLOGY

• Analysis of more than 650,000 tagged tweets concerning three different subjects: a) the death of Michael Jackson (MJ); b) the swine flu outbreak (SF); c) the 2010 Brazilian presidential election (BE), also divided into four sub-datasets: Dilma Roussef's supporters (BE-1); José Serra's supporters (BE-2); Dilma Roussef's opposers (BE-3); and José Serra's opposers (BE-4).

• Investigation of each gender users' choices in situations that admit linguistic variation – that is, the usage of different linguistic forms even when the semantic and the functional values are equivalent.

• As our datasets include users' given names, we assigned to each user a gender based on his/her name.
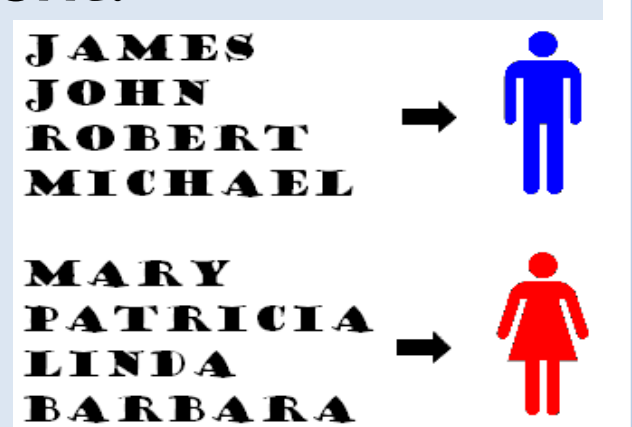
• Usage of z-scores acting as scaling factors so that comparisons between male and female users can take place using a common yard stick.

RESULTS

• We found that neutral hashtags correspond to great part of the corpus, but also that there is a significant presence of gendered tags.
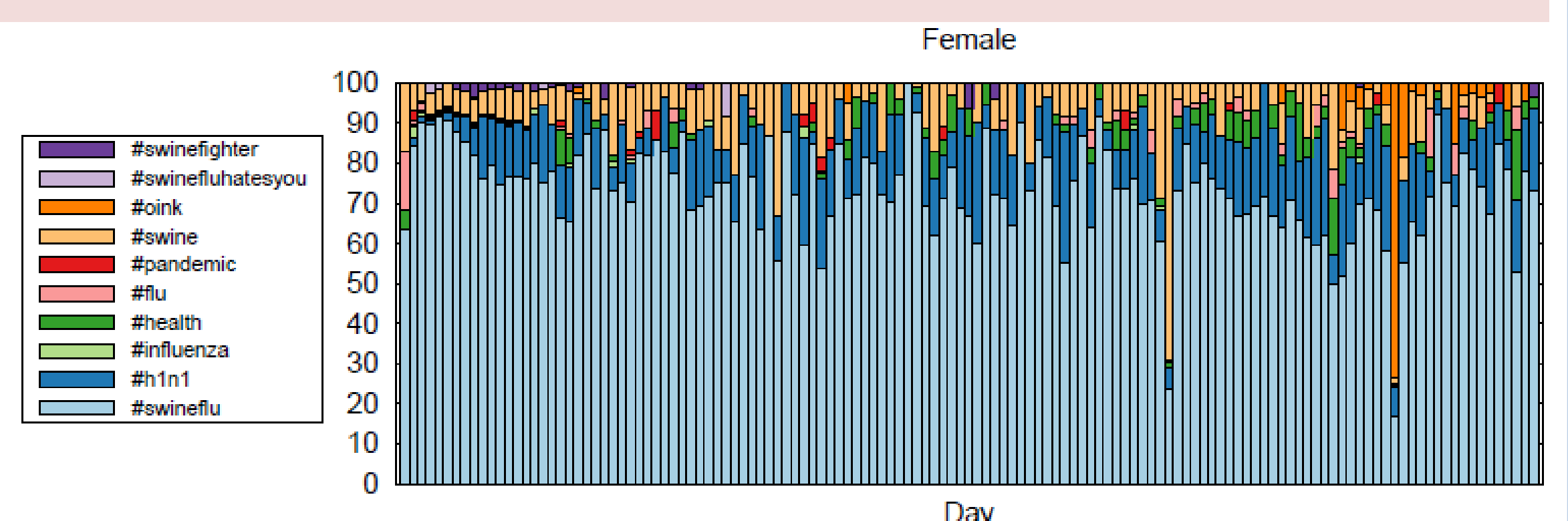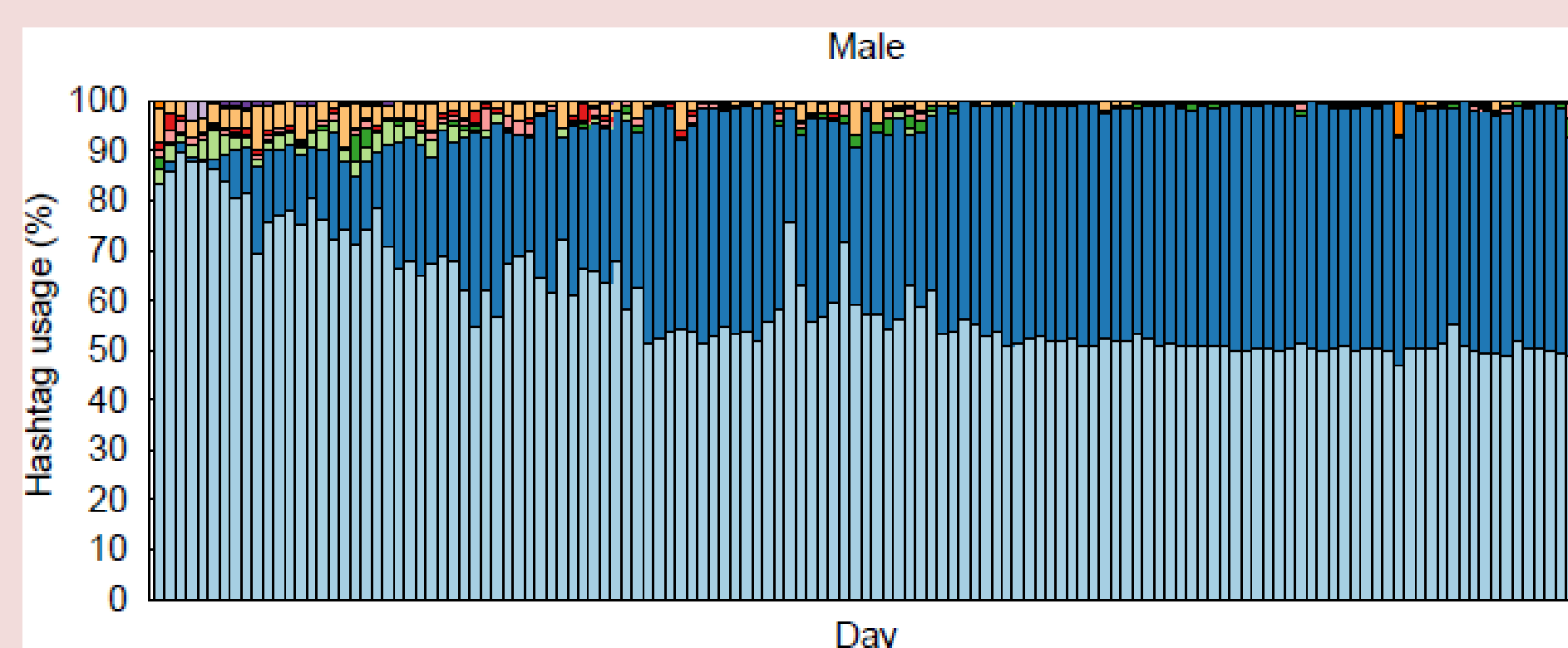
• For convenience, all z-scores presented in this study are associated to the female z-scores, which indicate the correspondent weight of the female usage. Therefore, positive z-scores will indicate a prevalence of female users and negative z-scores will indicate a prevalence of male users.

Standard vs. substandard forms

• We checked if the opposition between standard and substandard hashtags might be a factor that affects their acceptation by men and women.

• We consider as "standard forms" the most used hashtags, as they tend to be very transparent about the topic to which they refer. On the other hand, we consider the less common – generally more opaque – hashtags among those related to a given topic as the most innovative and "substandard forms".

• **Female users are more likely to use the most common hashtags, which we consider here as the standard forms, than male users.**

• **With the exception of the dataset SF, women also use more frequently the standard forms than the substandard ones.**

Dataset	Z-scores	
	Standard forms	Substandard forms
BE-1	0.974	-0.145
BE-2	0.450	-0.215
BE-3	1.024	-1.512
BE-4	0.885	0.031
MJ	1.467	-0.024
SF	0.002	0.079

Personal involvement vs. clear persuasion

• When analyzing the supporters of the candidates, we could distinguish some hashtags between two categories: those which included personal and direct user's involvement (written in singular 1st person, e.g. #votodilma/#votoserra ["I vote for Dilma/Serra"]), and those which contained pure voting instructions to the readers of the message produced (using the 3rd person imperative mood, expressing a command urging the audience to act a certain way, as in #vote13/#vote45 [13/45 = Dilma's/Serra's number]).

• Previous studies have found that men are more confident about their ability to persuade and that female managers, when attempting to convince a subordinate, rely more often on altruism than male managers.

• **"Personal involvement" strategy appears to be more common among female users. Male users, on the other hand, tend to be adopters of the "clear persuasion" strategy.**

Dataset	Z-scores	
	Personal involvement	Voting instruction
BE-1	0.601	-1.894
BE-2	1.477	-0.957

CONCLUSIONS AND FUTURE WORK

• This study provides evidence that, although the majority of the #hashtags seem to be neutral, some of them are, to some extent, gendered.
• We have also analyzed different categories of #hashtags and found that certain social roles occupied by each gender in offline communities are equally performed in online social networks and in content tagging.

• Future work will involve the improvement of the statistical analysis and the investigation of other aspects that might act as factors to influence users to employ particular tags.