

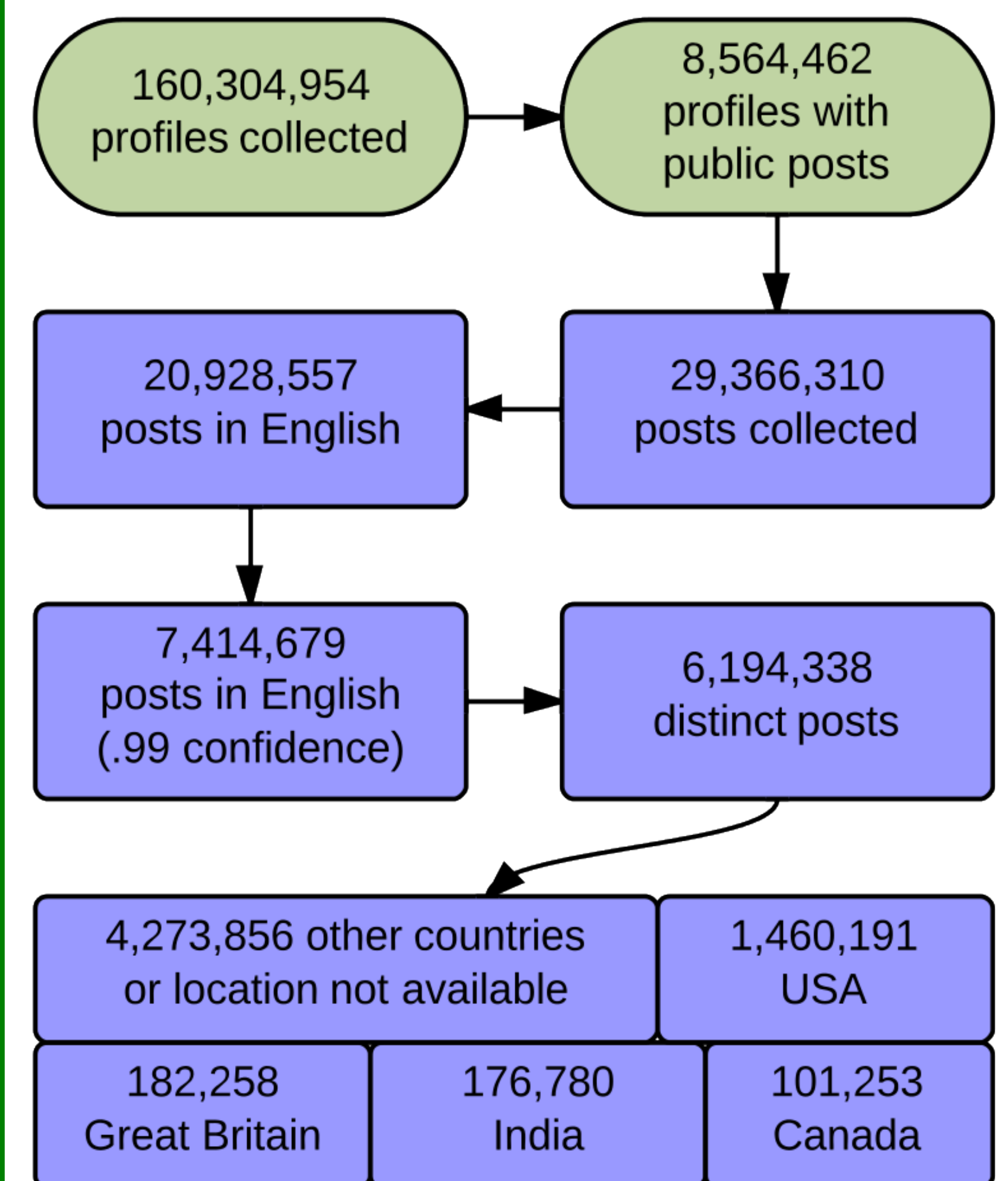
A Linguistic Characterization of Google+ Posts across Different Social Groups

Evandro Cunha, Gabriel Magno, Marcos André Gonçalves, César Cambraia, Virgilio Almeida
evandrocunha@dcc.ufmg.br, magno@dcc.ufmg.br
Universidade Federal de Minas Gerais, Brazil

Introduction

- Characterize Google+ status updates to better understand:
 - the textual genre 'post'
 - collective aspects of OSN users
- We consider about two million Google+ posts in order to evaluate linguistic elements among members of distinct social groups
- Applications in the development of improved interactive Web services:
 - detection of users' social characteristics based on textual analyses to:
 - find fake entities in social groups
 - personalize recommendation services

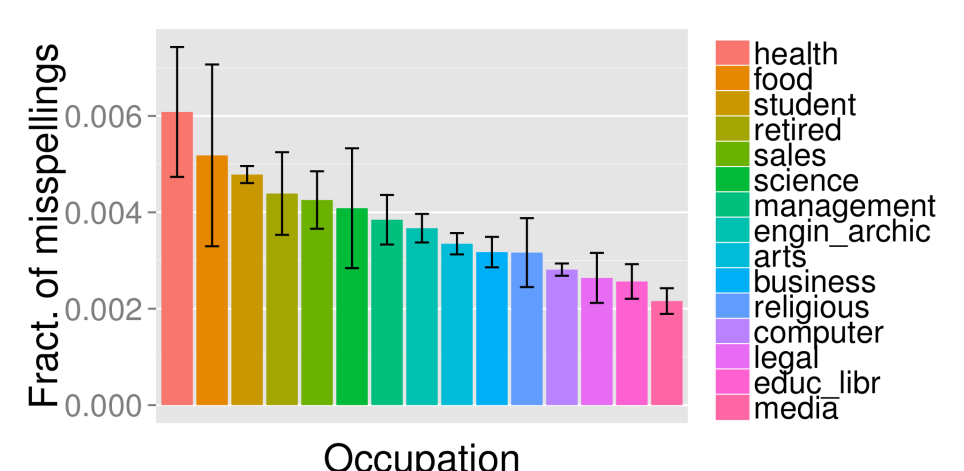
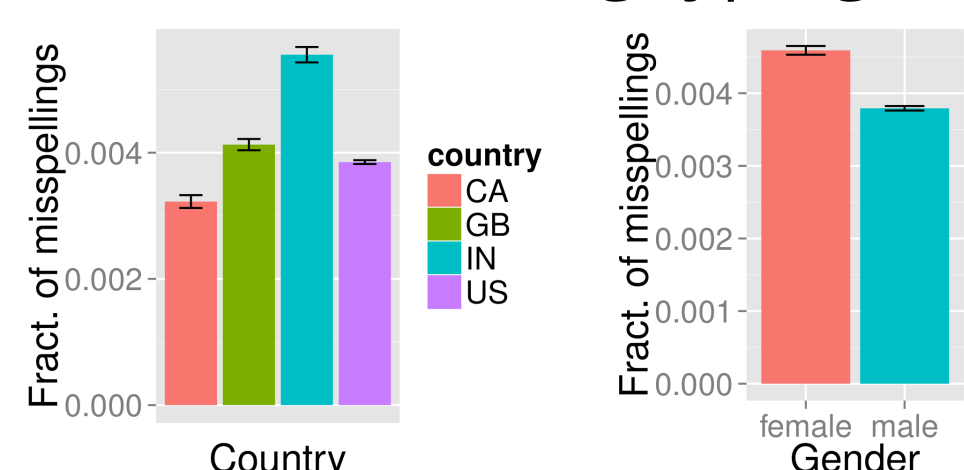
Dataset



Analyses and Results

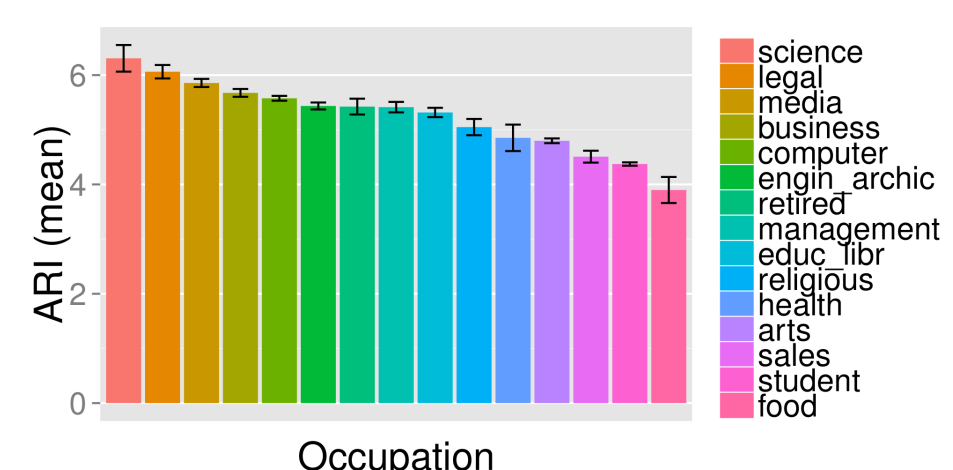
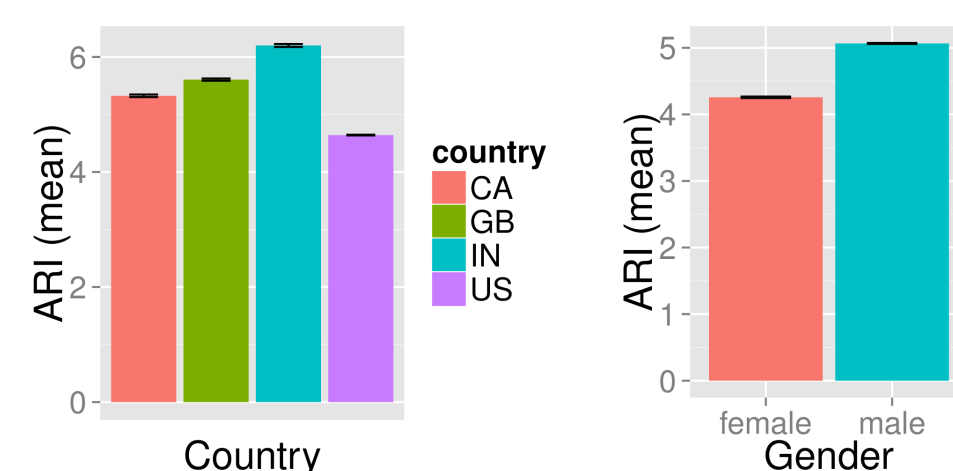
{ Misspellings }

- Fractions of misspellings per post
- Indicate low literacy levels or carelessness during typing



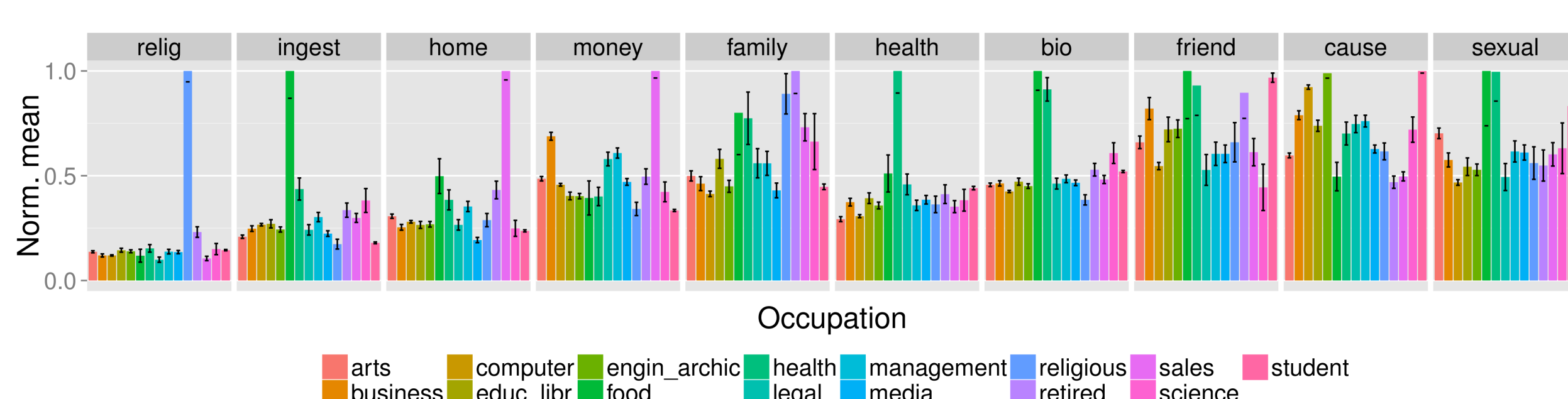
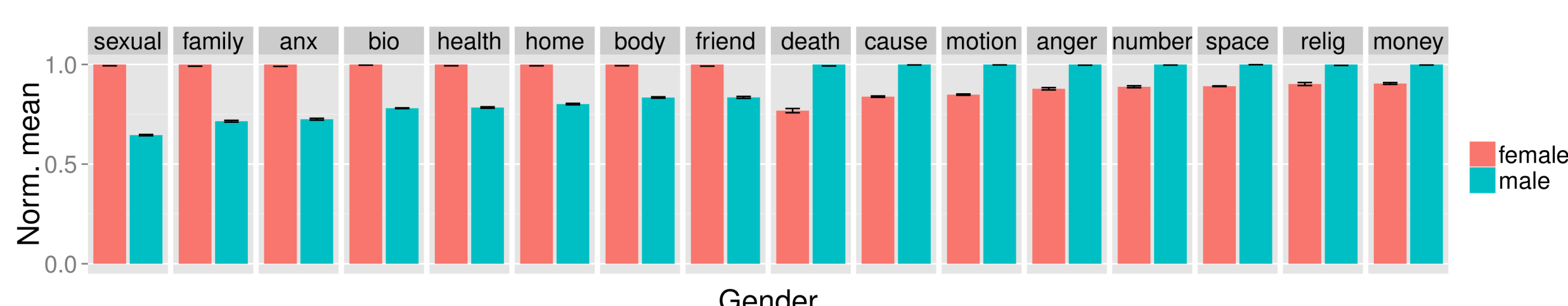
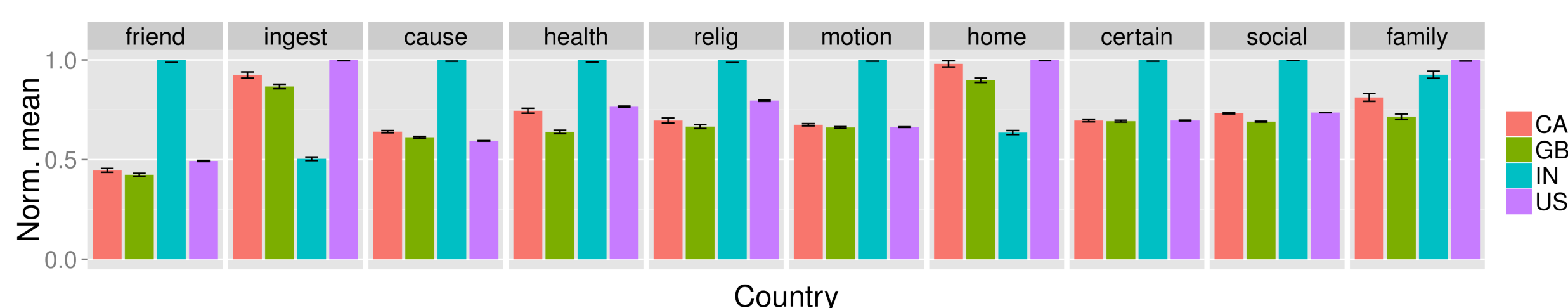
{ Readability and structural complexity }

- Quantify how simple a text is to be understood
- Higher scores indicate higher structural complexity, as they correspond to bigger words and sentences



{ Semantic categories of words }

- Reveal how groups perceive reality, showing what their main concerns are
- Language Inquiry and Word Count (LIWC) categories

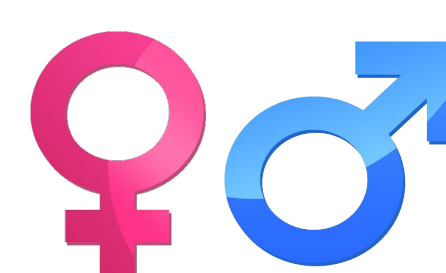


Social Groups

{ Countries }



{ Genders }



{ Occupations }

- Architect. and engin.
- Arts and design
- Business and financial
- Computer and math.
- Educ. and library
- Food prep.
- Healthcare
- Legal
- Management
- Media
- Religious
- Retired
- Sales
- Science
- Student

Conclusions

- Groups hold linguistic particularities:
 - different groups → different structures/vocabulary
 - women → social and familial relations
 - men → technical topics and achievements
 - tendency to keep a professional vocabulary
 - Google+ might be seen as a professional OSN
- Future work:
 - analysis of other relevant linguistic/social factors
 - e.g. topic of posts and educational level of users
 - compare these results with other popular OSNs