The effect of the author's sex and gender on the parameters of written texts with and without gender deception

Tatiana Litvinova, Voronezh State Pedagogical University
Pavel Seredin, Voronezh State University
George Mikros, National and Kapodistrian University of Athens

QUALICO-2018
Wroclaw, Poland, 4-8 July 2018

Gender in language

- Long history in sociolinguistics, many findings, although inconsistent

  Penelope Eckert, Sally McConnell-Ginet, Language and Gender (2nd edition, Cambridge University Press, 2013)

- Hot topic in ML, popular among PAN participants
- Many papers report high accuracies, but in cross-topic and cross-genre settings accuracy of prediction is shown to decline

Lessons from PAN RusProfiling Shared Task

- Training – Twitter, test – essays, Facebook, Twitter, reviews and gender-imitated texts
- Best results were for on Facebook, worst results on reviews.
- In case of gender-imitated texts, most systems failed. This is the most difficult scenario

- *Tatiana Litvinova*, *Francisco Rangel*, *Paolo Rosso*, *Pavel Seredin*, *Olga Litvinova,* **Overview** of the **RUSProfiling** PAN at FIRE Track on Cross-genre Gender Identification in Russian, http://ceur-ws.org/Vol-2036/

Why gender prediction from tweets is so hard?

- Nquen et al. (2014) claim that most research has so far ignored the fact that language use is related to the social identity of speakers, which may be different from their biological sex.
- **Gender or sex?**

Indeed, most studies, especially in ML, deal with the biological sex (the anatomy of an individual's reproductive system, and secondary sex characteristics) of the authors (**male/female**), not with gender, which can refer to either social roles based on the sex of the person (gender role) (**feminine/masculine**), because special tests are needed to detect level of feminity/masculinity.

So what do we really predict – biological sex or gender?

Stability (instability?) of linguistic features in idiolect

- Genre, topic of the text, mental state of the author, etc. – all these factors effect our language style
- Latest study by Dzogang, Lightman, Cristianini (2018) "Diurnal variations of psychometric indicators in Twitter content" has shown that even the time of day causes language changes, which associates with major changes in neural activity and hormonal levels
- Our study (Litvinova et. al., 2018, to appear) has shown that most of LIWC parameters which are widely used in authorship profiling studies are unstable in individual's idiolect.
- To date, there have been no special comprehensive studies which assess the level of stability of a wide range of linguistic parameters in an author's idiolect.

Possible style imitation

- People may change their writing style to hide their identity (for example, male authors pretend to be female)
- This is a problem of great practical importance
- Gender imitation was the most difficult scenario for classifiers developed by RusProfiling shared task participants
- Despite its obvious importance, the problem is understudied
- We still do not know which linguistic feature are manipulated in gender switching.

Gender Imitation Corpus

- N = 142, students of Russian universities
- Each respondent was asked to write 3 types of texts on the same topic:
  - ✓ Without deception;
  - ✓ With gender switching (imitation)
  - ✓ With style obfuscation

Rich metadata:

- ✓ Biological sex,
- ✓ age,
- ✓ handedness,
- ✓ psychological gender (IS – femininity/masculinity measure according to Bem Sex Role Inventory)

Freely available and expanding

Dataset (balanced)

For this study we used balanced dataset derived from Gender Imitation Corpus

- ✓ N = 60 (30 male, 30 female)
- ✓ 3 texts from each author
- ✓ Authors with different values of IS were chosen;
- ✓ All authors are native Russian speakers

Features

- High-frequent
- Context-independent
- From different linguistic levels
- Normalized by text length
- Moderately correlated with each other within each group

- POS features, especially function word, have shown to effectively discriminate male and female texts **(Koppel et al., 2002; Mikros 2013)**

Parameters

- 3 groups of parameters after selection

FW-based	QUITA	POS-based
Function words Pers Pron CONJ NOUN PRON *100 Most Frequent Words*	TTR R1 WritersView CurveLength R Max Token Length	ADV CONJ PART coef_VMvsSP speach_quality

Statistical analysis

- MANOVA/MANCOVA with follow-up DFA using SPSS
- 2 series of experiments:
- Separated MANOVAs for each one of the three stylometric groups with the biological sex and a text type (without deception, gender imitation, style obfuscation) as independent variables
- Separated MANCOVAs for each of the three stylometric groups with text type (without deception, gender imitation, style obfuscation) as a factor independent variable and psychologic gender as continuous independent variable (a covariate)

Results

- Gender,
- text type
- as factors

FW	QUITA	POS
Sex – significant, Pillai's = 2,488, p=0,033, partial Eta Squared =0.068 Type of text – n/s	**Sex – significant,** Pillai's = 2,411, p=0,038, partial Eta Squared =0.066 **Type of text – significant,** Pilai = 2,312, p=0,012, partial Eta Squared =0.066 **Joint – significant,** Pillai's = 2,666, p=0,004, partial Eta Squared =0.072	**Sex – significant,** Pillai's = 3,845, p=0,003, partial Eta Squared =0.103 Type of text – n/s

Results

- Text type as factor, psychological gender (IS) as covariate

FW	QUITA	POS
IS – n/s Type of text – n/s	IS – n/s, **Type of text – significant**, Pillai's = 2,323, p=0,012, partial Eta Squared =0.063	IS – n/s, **Type of text – n/s**

Results (DFA)

FW	QUITA	POS
SEX, Wilks' Lambda to test Equality of Group Means **FW** 0,955, p=0.004, **PRON_NOUN** 0,962, p=0,009 **100FREQ** 0,956, p=0,005 Test of Function – significant, p=0,031 Accuracy of the model is **63,3 %** (61,1 for female, 65,5 for male)	SEX, Wilks' Lambda to test Equality of Group Means **Only TTR,** 0,972, p=0,047 Test of Function – significant, p=0,039, Accuracy of the model is **60,6 %** (62,2 for female, 58,9 for male)	SEX, Wilks' Lambda to test Equality of Group Means **Only PART,** 0,925, p=0,000 Test of Function – significant, p=0,002 Accuracy of the model is **61,6 %** (62,2 for female, 60,9 for male)

Results (DFA)

QUITA
Type of text, Wilks' Lambda to test Equality of Group Means
TTR, 0,961, p=0,030, stand. coef. = -0,183
CurveLength, 0,955, p=0,017, stand. coef. = 0,545
Token 0,930, p=0,002, stand. coef. = 0,531
Test of Function – significant, p=0,011,
Accuracy of the model is **43,9 %**

Results

- There are linguistic features which are affected by authorial biological sex even in gender imitation and style obfuscation scenario
- There are statistically significant differences in function word based features in male and female texts even in gender imitation and style obfuscation scenario
- There are statistically significant differences in text frequency features in texts with/without gender imitation or style obfuscation

Function words usage

- **PRON_NOUN** stand. coef. = 0,732
- **100FREQ** stand. coef. = 0,579
- **FW** stand. coef. = 0,425
- All these parameters have higher values in male texts
- Male authors demonstrate the highest level of **PRON_NOUN, 100FREQ** and **FW** in text type 2 (gender imitation scenario)
- Female authors demonstrate the highest level of **PRON_NOUN and FW** in text type 3 (style obfuscation scenario)
- Although statistically non-significant, male authors also tend to use more personal pronouns in gender imitation scenario, while female authors use more conjunctions in this scenario

POS

- In gender imitation scenario, women use fewer adverbs, while men use more adverbs than in other types of texts
- In gender imitation scenario, both women and men use more nouns
- Only differences in numbers of particles are statistically significant across all types of texts: men use more particles, while reducing their number in gender imitation scenario (women use more particles in gender imitation scenario in comparison with other text types)

Relation to previous findings

- Our results are in contrast with previous findings based on texts without gender deception:

1) Female texts more “informal”, i.e. contains more pronouns and interjections (Rangel and Rosso, 2013)
2) Litvinova et. Al. (2017) have found that in male texts there are fewer most frequent Russian words, the majority of which are function words
3) Mikros (2013) revealed increased percentage of adverbs in male texts

Plausible explanation

- Respondents were aware of the full task, and could unconsciously change their style even while writing texts without stylistic deception
- Men and women have intuitive knowledge of characteristic features of their gender writing styles and manipulate them
- Men change their writing styles more dramatically than women

Text frequency features

- TTR is higher in female than in male texts in “truthful” condition contrary to what has been shown earlier (Litvinova, 2017), even more in gender imitation condition, and in style obfuscation scenario TTR is higher in male texts
- The problem with TTR is its dependence on text length
- CurveLength R which is less text length dependent are lower in texts with deception, with the lowest values in style obfuscation scenario
- Lower values of maximum token length is typical for texts with deception, with the lowest values in style obfuscation scenario

What to take on board

- One should be careful with POS and FW-based features as they are manipulated in gender imitation scenarios
- Text-independent indicators related to frequency structure of the text are worth investigating in the stylistic deception field
- While gathering corpora for gender imitation and – broadly – stylistic deception studies, it is important to allow respondents write “truthful” texts without knowing that they are to write deceptive texts

Future work

- Expanding our corpus using crowdsourcing platform
- Gender + age imitation
- Classification experiments using Stylo
- Cross-language perspective (Russian, Greek…)

Acknowledgment

- This work is supported by the grant no. 16-18-10050 from the Russian Science Foundation, project “Identifying the Gender and Age of Online Chatters Using Formal Parameters of their Texts”

Questions?

- Visit our website: http://en.rusprofilinglab.ru/
- Download Gender Imitation corpus: http://en.rusprofilinglab.ru/korpus-tekstov/gender-imitation-corpus/
- Create corpus in your language
- Contact us: centr_rus_yaz@mail.ru