

# An Examination of Language Use in Online Dating Profiles

Meenakshi Nagarajan, Knoesis, Wright State University, Dayton, OH  
 Marti A. Hearst, School of Information, Univ. of California, Berkeley, CA

## Overview

This work contributes to the study of self-presentation in online dating systems. The larger goal is to understand effects of self-presentation via free-text on perceived attractiveness.

### Motivating Question

What are the similarities and differences between how men and women self-present in online dating profiles?

## Data and Tools

To date, study of free-text components has been limited to few 10s of profiles. We extend study on a dataset of

- 500 male and 500 female profiles
- restricting reported ages between 18 and 60 and
- restricting to those looking for people of the opposite sex

Data was crawled from Yahoo! Personals, a popular, paid online dating site.

	Male	Female
Total No. of Words in Corpus	72,049	78,841
Avg. No. of Words per Profile	134	145

We used 80 features from **Linguistic Inquiry Word Count (LIWC)** spanning Linguistic, Psychological, Personal, Paralinguistic and Punctuation categories.

All statistical analysis were performed using **XLSTAT**, a statistical software for MSExcel

## Analysis – Part 1

**Exploratory Factor Analysis** to identify systematic co-occurrence patterns among LIWC features in profiles.

Used *factor analysis* and *'varimax rotation'* to delineate factors.

**Finding 1: Many similarities in overall factor structures between male and female profiles.**

Factors, loadings	Male Profiles	Female Profiles
F1, +ve loadings	Personal and 1st person singular and impersonal pronouns, common, auxiliary and present tense verbs	Personal and 1st person singular pronouns, common, auxiliary and present tense verbs
F1, -ve loadings	None	Comma (punctuation) and words longer than six letters
F2, +ve loadings	Affect, positive emotion, biological process and sexual words	Affect, positive emotion, biological process and sexual words
F2, -ve loadings	Relativity and space words	Relativity, space and time words
F3, +ve loadings	Impersonal pronouns, cognitive process, tentative and exclusion words	Impersonal pronouns, cognitive process, tentative, discrepancy and exclusion words
F3, -ve loadings	Relative words	Biological process and sexual words

All variable loadings in both datasets were greater than 0.4, with eigenvalues of 3.9, 2.6, 1.6 in the Male profiles and 4.7, 2.9, 2 in the Female profiles. The only crossloading variable was the 'impersonal pronouns' in the Male profiles with a maximum loading of 0.37.

### Finding 2: Tentative word usage

	Male	Female
% of all words used that were tentative	2.6%	2.2%
Mean no. of tentative words per profile	3.5	3.2
Std. dev of tentative words per profile	2.4	2.1

Tentative words are typically attributed to female discourse, suggesting that men might be temporizing their language in this setting.

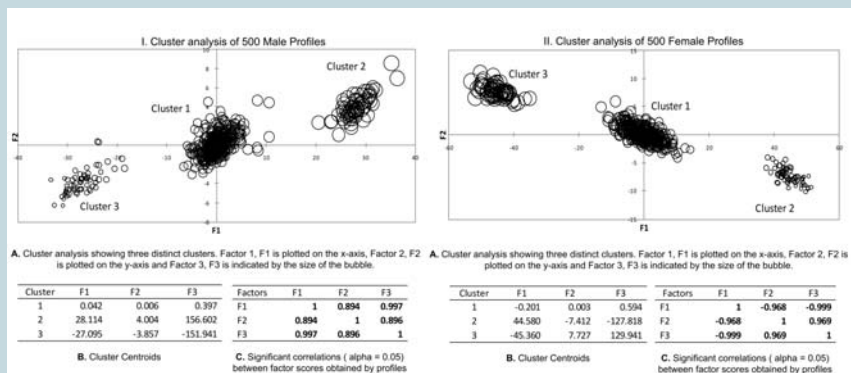
## Analysis – Part 2

Calculated factor scores for all profiles. High factor scores imply variables in profiles have high values in a factor pattern. A profile may load anywhere on the three factors.

Performed **Cluster Analysis** using profile factor scores as predictors. Profiles grouped in any cluster are intended to be maximally similar in their use of the features, while different clusters are maximally distinguished.

### Finding 3: Differences in male and female profile types

A k-means cluster analysis uncovered three distinct clusters for each dataset highlighting some contrasts between the sexes in their combinations of word types.



**Cluster 1:** 371 male and 363 female profiles showed moderate scores on all three factors.  
**Clusters 2 and 3:** showed an interesting contrast in the bundling of LIWC features.

	High use of	Moderate use of	Low use of
66 Male profiles with high scores on Factor 1 had	Affect, positive emotion, sexual and biological words; impersonal pronouns, cognitive process, tentative and exclusion words	CLUSTER 2	
		Sexual and biological words	Relativity words
70 Female profiles with high scores on Factor 1 had	Relativity words	Sexual and biological words	Affect, positive emotion words; impersonal pronouns, cognitive process, tentative, discrepancy and exclusion words
63 Male profiles with low scores on Factor 1 had	Relativity words	CLUSTER 3	
		Sexual, biological, affect, positive emotion words; impersonal pronouns, cognitive process, tentative, and exclusion words	Relativity words
67 Female profiles with low scores on Factor 1 had	Impersonal pronouns, cognitive process, tentative, discrepancy and exclusion words	Sexual, biological, affect, positive emotion	Relativity words

## Analysis – Part 3

We also conducted an analysis of the words used by the two genders.

**Finding 4: Gender similarities in word-usages across LIWC categories**

- For each LIWC category, the proportion of words used from that category versus all words in the corpus were comparable for male and female profiles.
- For almost all LIWC categories, 7 out of 10 high frequency words used by the gender groups were the same. High similarities were also seen in the usage of closed-class words (pronouns, prepositions etc.) and open-classed word categories (affect, emotion words, verb groups).
- Possible hypothesis:** Self-expression might tend toward attempting homophily in online dating profiles.