

9 July 2011

# The Role Of Linguistic Properties In Online Dating Communication – A Large-Scale Study Of Contact Initiation Messages

Valentin Schoendienst

*Humboldt-Universität zu Berlin*, [valentin@schoendienst.net](mailto:valentin@schoendienst.net)

Linh Dang-Xuan

*University of Münster*, [xuan@uni-muenster.de](mailto:xuan@uni-muenster.de)

ISBN: [978-1-86435-644-1]; Full paper

---

## Recommended Citation

Schoendienst, Valentin and Dang-Xuan, Linh, "The Role Of Linguistic Properties In Online Dating Communication – A Large-Scale Study Of Contact Initiation Messages" (2011). *PACIS 2011 Proceedings*. 169.  
<http://aisel.aisnet.org/pacis2011/169>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# THE ROLE OF LINGUISTIC PROPERTIES IN ONLINE DATING COMMUNICATION – A LARGE-SCALE STUDY OF CONTACT INITIATION MESSAGES

Valentin Schöndienst, Department of Information Systems, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany, valentin@schoendienst.net

Linh Dang-Xuan, Department of Information Systems, School of Business and Economics, University of Münster, Münster, Germany, linh.dang-xuan@uni-muenster.de

## Abstract

*For people who look for a partner, online dating largely increases the pool of potential mates. At the same time, users of online dating platforms have to cope with a large number of approaches and, therefore, need to choose selectively who they decide to engage in a conversation with. Especially, since the costs of rejection are low on online dating platforms, it is a common strategy to spam others with superficial approaches. With this in mind, and in the absence of nonverbal cues, targets base their decision of whether or not to respond to a message on (a) their impression of the sender's pictures, and (b) cues which they extract from the content of the message. The purpose of this study is to hypothesize on which linguistic properties of a message in computer-mediated communication may signal various qualities of its sender, to predict how those properties determine a target's decision of whether to respond or to ignore an initial message. Employing the Linguistic Inquiry and Word Count (LIWC) text analysis, relevant variables are operationalized from a corpus of 167,276 initial messages of an online dating platform. Regression analysis is performed in order to test the hypotheses. Results are discussed with respect to design implications for online dating platforms.*

*Keywords: Computer-mediated Communication, Online Dating, Text Analysis.*

# 1 INTRODUCTION

Online dating has long put behind what used to be considered the last hope for some desperate ones. In fact, mainstream picks up on online dating. In the U.S., 40 million, in China even 140 million, people throughout all ages and social classes seek a partner through the Internet.<sup>1</sup> Meanwhile, studies confirm preconceptions that people misrepresent themselves to attract partners. In their online dating profiles, men tend to lie about their age, height and income while women are likely to distort the truth about their weight, physical build and age (Hancock et al. 2007).

People seem to know this as they pay little attention to those variables. In fact, Fiore et al. (2008) show that fixed-choice variables such as age, height, political views, social setting, and ethnicity have no significant influence on how a perceiver evaluates the attractiveness of a person depicted in an online dating profile.<sup>2</sup> Instead, they find that an individual's attractiveness evaluation of a profile is determined by only two elements: (1) the shown picture, and (2) the free-text component of the profile.

Free-text is rich of implicit information (Tausczik & Pennebaker 2010) which provides important signals for potential mating partners. In this sense, some signals "give" meaning purposely, while additional unintended information is "given off" (Goffman 1959). From an evolution-theory perspective, signals are effective because they decrease the receiver's uncertainty regarding future behaviors of the signaler (Krebs & Dawkins 1984), i.e., the sender of a message in the present context. In an online dating environment, this is of particular importance as high social distances and the anonymity of the users make antisocial behavior more likely and false self-advertisement harder to reveal.

In face-to-face communication, nonverbal cues such as vocal intonation or gestures account for more of a receiver's perception of a sender's affect than the actual verbal content does (Burgoon et al. 1996). In online dating, as in computer-mediated communication in general, such nonverbal cues are absent. Here, linguistic traits of a message, with which a person approaches a target, can serve as signals of the sender's personality, abilities and qualities. Messages which are exchanged among users of an online dating platform provide recipients with both intended and unintended information about the sender. While it is easy to misrepresent explicit information such as age, weight, height and income, it is much harder to control what is given off "between the lines" of free-text. For example, eloquence may signal social status, but is difficult to fake for someone who is not articulate.

Recently, the emerging field of language psychology has produced fascinating results in revealing what people disclose about themselves through the words they use as they talk and write (e.g., identifying suicidal poets, Lightman et al. 2007; students suffering from depressions, Rude et al. 2004). Therefore, we analyze contact approaches, i.e., messages from individuals who had no previous interaction with the receiver of the message, with respect to how various linguistic properties determine the receiver's decision to respond or not. In other words, the present work investigates the research question: What makes a successful advance in online dating?

The results may provide insights into how analyzing people's written words can gain an understanding of which qualities people look for in a partner in an online dating environment. Further, the results may be used to augment current feedback-based matching algorithms (which tend to be fuelled with false information) with information which can be automatically extracted from the messages people exchange among each other. The contribution of this work is meant to be a step

---

<sup>1</sup> See van Grove (2010) for an overview of relevant statistics.

<sup>2</sup> Hitsch et al. (2006) show the significance of some fixed-choice variables in determining online dating outcome. However, the studied online dating profiles did not include free-text elements which, in the light of Fiore et al.'s (2008) findings, may have had made those elements insignificant.

away from feedback-based matching algorithms and towards a match-making approach which mines people's real behavior.

To this end, we proceed as follows. In the following section, we derive hypotheses from the literature on interpersonal relationships and linguistic text analysis. We then perform regression analysis to empirically test the formulated hypotheses using a data set of 167,276 messages from an online dating platform. Finally, we conclude by discussing the results with respect to implications for the design of online dating platforms.

## 2 THEORETICAL BACKGROUND

The life-cycle of heterosexual adult romantic relationships follows five stages: acquaintance, build-up, continuation, deterioration and termination (Levinger 1983). Much research has been devoted to each of these phases and their transition phases. Relationship initiation has received scholarly interest especially with respect to the strategies that people use to attain sexual intimacy. Also, research on interpersonal attraction has investigated strategies which males and females apply to increase their appeal. This work looks at the other end of the trade: the ultimate decision of the target to respond to an approach.

We draw on findings from evolutionary psychology as well as the broad field of interpersonal relationships in order to hypothesize how various personal traits as derived from linguistic dimensions may translate into a target's decision whether to respond to an attempt or not. However, since research on mate choice emerged from the study of non-human animals, assessments of mate qualities often do not consider psychological traits such as personality and attitudes.

Therefore, it has been criticized that despite a broad body of research on physical attractiveness (skin tone, face symmetry etc.), "little is known precisely which characteristics in potential mates are valued by human males and females" (Buss 1989, p. 1; Buss 1985; Thiessen & Gregg 1980). However, research on intersexual selection has produced some insights regarding the preferences for particular qualities in potential mates. Studies from this research line, generally referred to as mate choice, provide an understanding of factors which determine people's initial dating decision. In this light, the following section identifies relevant factors from the literature. For each factor, the literature is reviewed with respect to how linguistic traits of a message can serve as an indicator for the possession of particular qualities by the sender.

Differences in interpersonal relationships due to biological sex (i.e., male versus female) in conjunction with socialization and learning experiences lead to psychological differences between males and females in human mate preferences (Buss 1989; Brehm 1992). Therefore, this study also examines gender differences in the perception of mate qualities through online dating messages.

## 3 HYPOTHESES

The psychological meanings of words have been investigated by various scholars.<sup>3</sup> Among others, linguistic traits include the degree of self-reference in a message, the extent to which messages contain words which refer to social processes or leisure as well as the use of words related to positive versus negative emotions. In the following, we review the literature with respect to how the usage of words of particular categories in a message may signal various personality traits and formulate hypotheses on the effects on a target's decision to respond to a message or not. Further, recent research has also examined how the use of certain word categories triggers responses in non-romantic settings. Findings from this research strand are also included in the formulation of the hypotheses.

Given the importance of physical attractiveness in relationship initiation (e.g., Walster et al. 1966), the role of physical attractiveness of both the sender as well as the target is also included in the analysis.

---

<sup>3</sup> See Tausczik and Pennebaker (2010) for an overview.

### 3.1 Physical Attractiveness

It is both intuitive and a consistent research finding on interpersonal attraction that physical attractiveness is the most important factor in determining people's initial dating decision. It applies to both males and females that the more physically attractive they find someone, the more likely they are to interact with and date a person (Walster et al. 1966; Buss & Barnes 1986; Feingold 1991; Regan & Berscheid 1997). Even though this work's focus is on the role of linguistic traits of messages, physical attractiveness as a major factor of people's dating rationale cannot be neglected from the analysis. Meanwhile, studies show that men put a stronger focus on physical attractiveness while women weight the importance of other factors higher.<sup>4</sup> This leads to our first hypothesis:

- H1: The more attractive the sender of an initial message is, the more likely a target is to respond. This applies to both (a) male and (b) female targets. However, (c) the attractiveness of the sender is more important for men than women.

On average, individuals end up with partners of similar attractiveness (i.e., assortative mating, Buston & Emlen 2003; Kowner 1995; Little et al. 2001; Todd et al. 2007). While this means that attractive people prefer other attractive people, it does not mean that less attractive people find more attractive people less appealing. In other words, an individual's own physical attractiveness does not appear to affect the perception of other people's attractiveness (Lee et al. 2008). Consequently, less attractive individuals are much more attracted to good-looking people than to other unattractive people. Therefore, attractive people receive more messages (Hitsch et al. 2006). It follows that attractive people have more choice which means the chances for each individual message to be answered are lower. Therefore, we hypothesize:

- H2: The more attractive a receiver of an initial message is, the less likely (a) he or (b) she is to respond.

### 3.2 Word Count as a Cue for Communicativeness

Studies indicate that for women talking is important for the maintenance of a relationship while men assign only little importance to talking for the maintenance of a romantic relationship (e.g., Riessman 2002). Women prefer a more talkative partner, while men prefer the opposite. The length of the first message a target receives from a sender can serve as an indicator of how talkative the sender might be in a relationship. Therefore, we expect women to be more likely to respond to longer messages while the opposite is true for men (i.e., less likely to respond to longer messages).

- H3: (a) The longer the initial message is, the less likely men are to respond while (b) women are more likely to respond to longer initial messages.

### 3.3 Usage of Self-references as a Cue for Depressive Symptoms

The use of self-reference words such as "me," "myself" and "I" indicates attentional allocation on the self and has been linked to various personal traits. In particular, increased usage of self-references has been found to be linked to depressive symptoms (e.g., Rude et al. 2004). Meanwhile, the articulation of depressive symptoms has been reported to lead to social rejection (Coyne 1976). Therefore, it is hypothesized that both male and female targets are less likely to respond to an initial message when it contains more self-references.

- H4: Usage of self-references in an initial message has a negative effect on the likelihood of triggering a response of a (a) male or (b) female target.

---

<sup>4</sup> See Feingold (1991) for an overview.

### **3.4 Usage of You-references as a Cue for Interest in the Target**

The use of personal pronouns such as “you” and “yours” indicates attentional allocation on the target. While people who are experiencing physical or emotional pain tend to have attention drawn to themselves (Rude et al. 2004), addressing the target directly may signal interest. This may lead to increased liking of the sender (e.g., Berscheid et al. 1976) and, consequently, higher likelihood to respond. From a theory perspective, there is no indication for gender differences regarding such an effect. It is therefore hypothesized that:

- H5: Addressing a target directly in an initial message makes (a) him or (b) her more likely to respond.

### **3.5 Usage of Social-processes Words as a Cue for Social Support and Leisure Words as a Cue for a Lack of Care-taking Ability**

Evolutionary theory suggests that in species with male parental investments including humans (Alexander & Noonan 1979), “females should seek to mate with males who have the ability and willingness to provide resources related to parental investment such as food, shelter, territory and protection” (Buss 1989, p. 2; Trivers 1972). Usage of words which refer to social processes of various kinds have been found to be linked to social support (e.g., Owen et al. 2003; Rellini & Meston 2007). Therefore, we expect the usage of words which are associated with social processes to increase the likelihood of receiving a response for men who approach women.

From an evolution-theory perspective, women’s mate choice is determined by their anticipation of the male’s willingness and ability to invest care and resources into provisioning them and their offspring (e.g., Nisbet 1973). If this is the case, men’s usage of words which are associated with leisure (e.g., “movie”) may indicate a lack of care-taking ability and is, therefore, hypothesized to decrease the chances of getting a reply. The reviewed literature does not indicate that this hypothesis could be equally applicable to men’s preferences in women. Therefore, we propose the following hypotheses:

- H6: (a) For women, usage of words which are associated with social processes does not influence their chances of receiving a response from a man. (b) When men use words in initial messages which are associated with social processes, women are more likely to respond to them.
- H7: (a) For women, usage of words which are associated with leisure does not influence their chances of receiving a response from a man. (b) When men use words in initial messages which are associated with leisure, women are less likely to respond to them.

### **3.6 Usage of Sexual Words as a Cue for Interest in Sexual Intimacy**

Men are more eager for sex than women (e.g., Clark & Hatfield 1989) and more likely to respond positively to flirtatious behavior (e.g., Abrahams 1994). Frisby et al. (2010) show that women do not respond positively when men flirt for sexual motives while men’s attraction to women increases significantly when women do so. Therefore, we expect the usage of words which are associated with sexual processes (e.g., “horny”) to have a positive effect on the chances to get a response when women approach men, but no significant effect when men approach women:

- H8: The more an initial message contains words which refer to sexual processes, the more likely a (a) male target is to respond. (b) Such effect, however, does not apply to female targets.

### **3.7 Positive Emotions and Negative Emotions as a Cue for Emotional Intimacy**

Reis and Shaver’s (1988) model of interpersonal intimacy promotes that people are seeking emotional closeness in their romantic relationships. Clark et al. (1999) find that the most frequently used behaviors in romantic relationship initiation are those that promote such emotional intimacy. Further, results from studies of online interactions in non-romantic setting (Joyce & Kraut 2006; Huffaker

2010) show that both negative and positive affect of messages can trigger feedback and involvement. Gender differences regarding the effect of emotion articulation are not indicated by the reviewed literature. It is therefore hypothesized that:

- H9: The more an initial message articulates positive emotion, the more likely a (a) male or (b) female target is to respond.
- H10: The more an initial message articulates negative emotion, the less likely a (a) male or (b) female target is to respond.

## 4 METHODOLOGY

### 4.1 Data and Variables

For the empirical test of our hypotheses, we employed two data sets provided by a large Australian online dating platform. The portal allows members to create user profiles and to send messages to other members. It is prerequisite for members to upload at least one picture of themselves. The authenticity of the pictures is verified manually by the provider. Each member can rate other members' physical attractiveness on a scale from 1 to 10. The attractiveness evaluation of each member is shown along with the user name and picture(s) on his or her profile page. After we were provided with the data set, the provider has implemented a feature which allows for the provision of further textual information such as hobbies and hair-color.

The first data set contains 1,002,555 messages exchanged by the members in the period from March 13, 2009 to July 22, 2010. The second data set contains the attractiveness evaluations of members. Based on the first data set, we extracted all messages which are considered initial messages sent to target persons, i.e., all messages which were sent without any prior message exchange between the sender and the receiver. As a result, we obtained a data set which contains - after excluding contact attempts among members of the same gender (i.e., among homosexual members) - a total of 167,276 initial messages sent by 3,657 distinct members. We then checked whether or not the initial message has triggered another message as a reply from the target and constructed a corresponding binary decision variable.

The attractiveness-ratings data set contains members' ratings of other members' attractiveness. In total, 89,785 ratings were made by 4,399 distinct members; each rating represents one member's assessment of another member's attractiveness after viewing the other member's pictures. For each member, we then calculated the average of his or her attractiveness ratings (i.e., the average of the ratings the member received from other members).<sup>5</sup>

We used the Linguistic Inquiry and Word Count (LIWC) Software (Pennebaker et al. 2006) to analyze initial messages for various linguistic traits. LIWC is a text-analysis software program that places words from a text file into categories based on a series of built-in dictionaries. These dictionaries have over 4,500 words and word stems containing a total of 80 categories into which words may fit. These categories include descriptive dimensions (e.g., total number of words in text, average number of words per sentence), linguistic dimensions (e.g., percentage of words in text that are pronouns or verbs), dimensions of psychological constructs (e.g., affect words, cognition words),

---

<sup>5</sup> Note that as there are also within-gender ratings in the attractiveness-ratings data set, there might be potential biases when ratings were made by homosexuals. We controlled for these biases by checking whether there has been a contact attempt (in terms of an initial message) between the rater and the target additionally. If this was the case, we assumed homosexuality and excluded the corresponding rating from our calculation of the average of the attractiveness ratings for the corresponding target. Also note that the ratings and reply decisions in the data sets were based on members' exposure to the photos of other members and hence were not colored by any face-to-face interactions between members and their contact targets.

dimensions of personal concerns (e.g., leisure, work), paralinguistic dimensions (e.g., fillers, assent), and punctuation.

The relevant LIWC categories for our analysis include “word count,” “first-person singular personal pronoun,” “second-person personal pronoun,” “social processes,” “leisure,” “sexual,” “positive emotion” and “negative emotion.”

In sum, the following variables were constructed for the empirical analysis:

- reply decision (binary): *REPLY*
- sender’s attractiveness (average attractiveness of the recipient of the message rated by other members): *SENDERATTRACT*
- recipient’s attractiveness (average attractiveness of the sender of the message rated by other members): *RECIPIENTATTRACT*
- LIWC categories: *WORDCOUNT, I, YOU, SOCIAL, LEISURE, SEXUAL, POSEMO, NEGEMO*

## 4.2 Analysis Method

We applied regression techniques to examine whether sender’s and recipient’s attractiveness as well as different linguistic dimensions affect individuals’ decision whether or not to reply to initial messages. We fitted a random-effects logit model to account for the binary dependent variable. The random-effects specification was chosen to control for targets’ heterogeneity (Baltagi 2008). For the purpose of comparison, we split our main sample into two sub-samples with the first one comprising only female-to-male (i.e., initial messages sent by females) approaches and the other only male-to-female (i.e., initial messages sent by males) advances.

# 5 EMPIRICAL RESULTS

## 5.1 Descriptive Statistics

Our data set comprises a total of 167,276 messages sent by 3,657 distinct members. Much more messages were sent by men than women (78% males and 22% females), which is in line with previous findings from the literature that men’s verbal communication is often more direct or overt than women’s (e.g., Berger & Bell 1988; Greer & Buss 1994). That is, men are more willing (e.g., Green & Sandos 1983) and more likely to initiate relationships than women, often by verbally requesting dates (e.g., Berger 1988; Kelley & Rolker-Dolinsky 1987).

## 5.2 Hypotheses Testing

Table 1 (Table 2) summarizes the results of analyzing the relationship between male (female) members’ decisions of whether or not to reply to an initial message from a female (male) member and a set of possible predictors. Under H1, we hypothesize that the more attractive the sender of a message, the more likely (a) he or (b) she will get a response to their message, and that (c) the attractiveness of the sender is more important for male than female recipients. Indeed, results from the baseline regression (Model 1) in Table 1 and 2 indicate that the likelihood of a member’s feedback to an approach after seeing his or her picture was positively predicted by the sender’s attractiveness (**H1a and H1b supported**). Moreover, the coefficient of *SENDERATTRACT* in Table 1 ( $b = 0.13, p < 0.001$ ) is higher than that in Table 2 ( $b = 0.06, p < 0.001$ ) suggesting that an increase in sender’s attractiveness would lead to a higher probability of a reply for men than women. **H1c** is therefore also **supported** by the data. This result is also consistent with previous findings showing that men put a stronger focus on physical attractiveness, while women weight the importance of other factors higher.

H2 predicts a negative relationship between the likelihood of a member's decision to reply to a contact attempt and (a) his or (b) her own attractiveness. However, H2 finds only partial support by the empirical results (Model 1). More specifically, while the coefficient of *RECIPIENTATTRACT* in Table 2 is significantly negative ( $b = -0.08, p < 0.05$ ), the same coefficient in Table 1 is significantly positive ( $b = 0.10, p < 0.001$ ). Our results suggest that the more attractive a woman is, the less likely she will respond to contact attempt (**H2b supported**). On the other hand, the more attractive a man is, the more likely he will make a positive reply decision (**H2a rejected**). This significant difference between men and women is surprising and deserves further investigation.

Under H3, we hypothesize that (a) the longer the initial message is, the less likely men are to respond while (b) women are more likely to respond to longer initial messages. We find strong **support for H3b** as *WORDCOUNT* is significantly positively related to the probability of a reply ( $b = 0.01, p < 0.05$ ; see Model 2, Table 2) indicating that women are prone to reply to longer initial messages. In contrast, we find that the length of a message does significantly affect men's reply decision in the opposite way, i.e., they are less likely to respond to longer messages ( $b = -0.01, p < 0.05$ ; see Model 2, Table 1). Therefore, **H3a is supported**. This finding indicates a fundamental gender difference regarding the effect of cues for communicativeness.

As predicted by H4, results show that self-reference in initial messages indeed tends to lower recipient's propensity to reply. Both coefficients of *I* in Model 3, Table 1 ( $b = -0.09, p < 0.01$ ) and 2 ( $b = -0.03, p < 0.01$ ) are negative and highly significant. The effect is even stronger in case of reply decision by men/women. Hence, **H4a and H4b are both confirmed**.

Under H5, we expect that as people like to receive attention, addressing a target directly by using more second-person personal pronouns such as "you" would make the recipient more likely to respond to the message. Results regarding the coefficient of *YOU* in Model 4, Table 1 ( $b = 0.10, p < 0.01$ ) and 2 ( $b = 0.08, p < 0.001$ ) clearly **confirm H5a and H5b** suggesting that usage of you-references in messages triggers more replies from both male and female targets.

To examine whether using more words which refer to social processes in initial messages increases the likelihood of a female target's response as predicted by H6b, we look at the coefficient of *SOCIAL* displayed in Model 5, Table 2. The coefficient is positive and statistically significant ( $b = 0.04, p < 0.001$ ) indicating that the more the sender of message uses words which refer to social processes, the more likely the female recipient is to respond. **H6b is thus supported**. However, **H6a is not confirmed** as men also tend to reply more often to messages containing social-processes words ( $b = 0.04, p < 0.05$ ; see Model 5, Table 1).

Under H7b, we expect that the more an initial message contains words which refer to leisure, the less likely a female target is to respond. Our results clearly **support H7b**. As shown in Model 6, Table 2, *LEISURE* is found to be statistically negatively associated with the likelihood of a reply by female targets ( $b = -0.15, p < 0.01$ ). On the other hand, the same effect applies male targets, i.e., men are also less likely to reply to messages containing leisure words ( $b = -0.22, p < 0.01$ ; see Model 6, Table 1). **H7a is therefore rejected**.

H8 predicts differences between males and females when they are confronted with initial messages containing words which refer to sexual processes. More specifically, men are more likely to respond to such messages (H8a) while this effect does not hold for women (H8b). In fact, we find **support for H8a** as the coefficient of *SEXUAL* is positive and statistically significant ( $b = 0.16, p < 0.01$ ; see Model 7, Table 1). **H9b is also supported** as the coefficient of *SEXUAL* is insignificant ( $b = -0.01, p = 0.68$ ; see Model 7, Table 2). This suggests that the occurrence of words which refer to sexual processes does not have an impact on female targets' reply decision.

Regarding affective dimensions of initial messages, we find only partial support for H9 which suggests that the more an initial message articulates positive emotion, the more likely a (a) male or (b) female target is to respond. More specifically, we find no significant relationship between *POSEMO* and the likelihood of a male target's reply ( $b = -0.05, p = 0.20$ , see Model 8, Table 1). That is, **H9a is rejected**. However, as shown in Table 2, *POSEMO* is significantly positively related to the probability of response by female targets ( $b = 0.02, p < 0.01$ , see Model 8, Table 2) suggesting that

women are more likely to respond to initial messages which articulate positive emotions (**H9b confirmed**).

Predictor	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>SENDERATTRACT</i>	0.13*** (4.84)	0.14*** (4.98)	0.15*** (5.04)	0.13*** (4.67)	0.13*** (4.73)	0.13*** (4.92)	0.13*** (4.87)	0.13*** (4.87)	0.13*** (4.92)
<i>RECIPIENTATTRACT</i>	0.10*** (3.65)	0.10*** (3.67)	0.10*** (3.67)	0.10*** (3.61)	0.10*** (3.60)	0.10*** (3.65)	0.10*** (3.63)	0.10*** (3.63)	0.10*** (3.71)
<i>WORDCOUNT</i>		-0.01* (-2.23)							
<i>I</i>			-0.09** (-2.55)						
<i>YOU</i>				0.10** (2.56)					
<i>SOCIAL</i>					0.04* (2.02)				
<i>LEISURE</i>						-0.22** (-2.87)			
<i>SEXUAL</i>							0.16** (2.58)		
<i>POSEMO</i>								-0.05 (-1.80)	
<i>NEGEMO</i>									-0.07 (-1.18)
Log likelihood	-5,893	-5,791	-5,787	-5,790	-5,792	-5,798	-5,799	-5,790	-5,789
Pseudo-R <sup>2</sup>	18.18%	20.46%	20.52%	20.48%	20.45%	20.36%	20.34%	20.48%	20.49%

Note that the table reports regression coefficients, with z-statistics in parentheses. This data set included 12,859 decisions of 1,883 male members. \*, \*\*, and \*\*\* indicate significance level at 5%, 1%, and 0.1%, respectively.

Table 1. Results of Random-Effects Logistic Regressions Predicting Reply Decision of Male Targets.

Predictor	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>SENDERATTRACT</i>	0.06*** (12.93)	0.06*** (12.98)	0.06*** (12.86)	0.06*** (12.83)	0.06*** (12.97)	0.06*** (13.00)	0.06*** (12.92)	0.06*** (12.89)	0.06*** (12.73)
<i>RECIPIENTATTRACT</i>	-0.08* (-2.02)	-0.08* (-2.04)	-0.08* (-2.02)	-0.09* (-2.10)	-0.09* (-2.09)	-0.09* (-2.05)	-0.08* (-2.02)	-0.08* (-2.02)	-0.08* (-2.01)
<i>WORDCOUNT</i>		0.01* (2.38)							
<i>I</i>			-0.03** (-2.88)						
<i>YOU</i>				0.08*** (9.65)					
<i>SOCIAL</i>					0.04*** (8.13)				
<i>LEISURE</i>						-0.15** (-3.06)			
<i>SEXUAL</i>							-0.01 (-0.99)		
<i>POSEMO</i>								0.02** (3.02)	
<i>NEGEMO</i>									-0.01 (-0.78)
Log likelihood	-59,840	-58,837	-58,831	-58,795	-58,808	-58,836	-58,840	-58,845	-58,806
Pseudo-R <sup>2</sup>	14.09%	15.75%	15.76%	15.82%	15.80%	15.75%	15.74%	15.73%	15.79%

Note that the table reports regression coefficients, with z-statistics in parentheses. This data set included 154,417 decisions of 1,416 female members. \*, \*\*, and \*\*\* indicate significance level at 5%, 1%, and 0.1%, respectively.

Table 2. Results of Random-Effects Logistic Regressions Predicting Reply Decision of Female Targets.

Under H10, we hypothesize a negative relationship between articulating negative emotions in initial messages and the likelihood of a positive reply decision by the target. However, our results do not corroborate such prediction with respect to both male and female recipients. The coefficients of *NEGEMO* in Model 9, Table 1 ( $b = -0.07$ ,  $p = 0.36$ ) and 2 ( $b = -0.01$ ,  $p = 0.60$ ) are both statically insignificant. Both **H10a** and **H10b** are therefore rejected.

## 6 CONCLUSION

The tremendous growth of the online dating industry (van Grove 2010) has led to the possibility for people to initiate contact with potential partners at a scale that is unprecedented in human history. Returning to our original research question, we can definitively state that linguistic properties of initial messages affect a target's decision of whether or not to respond to an advance. This shows that, in online dating, it is not only a person's physical appearance that decides over the success of an approach. The goal of this research was to determine which linguistic dimensions exactly determine people's decision to engage in a conversation. More specifically, it was the aim of this work to provide empirically valid insights into how particular word categories which people use in an initial message increase or decrease their chances of receiving a reply. Also, it was the purpose of the study to analyze how gender affects responsiveness to linguistic traits in initial messages on an online dating platform.

Here, a couple of interesting insights emerged from the results of the regression analyses. These allow us to derive a few implications for both the use and design of online dating platforms. Both females and males are less likely to respond to an initial message if the sender uses more self-references. People seem to associate personality traits (e.g., depressive symptoms; Rude et al. 2004) with such behavior which discourages them to further engage in conversations. Also, it is a hint that in online dating people are less interested in (potentially false) self-descriptions of individuals (in line with Fiore et al.'s (2008) findings). Instead, people seem to prefer to have attention drawn to them as messages containing second-person personal pronouns are more likely to be replied. This implies that in order to increase their chances of receiving a reply people should avoid self-reference. Rather, they should give the target more attention by addressing it directly.

As expected, female targets are more likely to respond to lengthier messages while men are less likely to reply to such messages. This gender difference suggests that men should invest more time to articulate longer messages while women should keep the communication shorter if they want to be more successful in receiving feedback.

We assumed usage of words which refer to social processes to signal social support (which would increase the likelihood of a target to respond) while words which refer to leisure would signal a lack of care-taking ability (decreasing the likelihood). Interestingly, these effects were significant for both males and females. Of course, these effects are most likely due to subconscious evaluation processes by a target. Following up on these findings, future research should investigate with more granularity how preferences in mate choice can be mined based on people's responsiveness to various linguistic properties. These findings are important because not only do people misrepresent themselves in online dating, also they fail to adequately state their own preferences (Hitsch et al. 2006).

Consistent with previous findings from the literature (e.g., Abrahams 1994; Frisby et al. 2010), men are more likely to respond to initial messages containing words which refer to sexual processes. This implies that women can allow themselves to be more flirtatious in their language when approaching a target. Men, however, do not increase their chances of getting a response when they "dirty-talk." Rather, the results show that expressing emotional state in initial messages, particularly positive emotions, increase their chances.

Regarding physical attractiveness, it is straightforward that the attractiveness of the sender positively affects a target's decision to reply to an initial message. However, with respect to the physical attractiveness of a target a surprising gender difference was found. As expected, more attractive women are less likely to respond to an approach, but the physical attractiveness of men turned out to be positively linked to the likelihood of a response of them to an initial message. This finding leads to

the counter-intuitive implication that women have better chances with more attractive men. This may be because those men receive less messages overall as women may not even dare to contact them. In any event, this finding deserves further investigation.

Previous research pointed at the relevance of free-text elements in determining online dating outcome (Fiore et al. 2008). According to the authors' best knowledge, the present study is the first to provide detailed insights into how free-text can be mined in order to predict online dating outcomes. It has been argued that relying on revealed rather than stated preferences might yield more reliable results for certain dimensions of mate choice. This is important because to date most online dating systems ask people to explicitly state information. As a design implication, providers of online dating platforms are advised to put more emphasis on providing members with means to articulate themselves on a free-text basis. Not only do the results show that people derive important cues from free-text. Also, the results indicate that, as the research in this field continues, match making could one day rely entirely on (1) mined preferences and (2) mined characteristics of mate-seeking individuals. This will eliminate the inconvenience for people to describe themselves and their preferences in online dating profiles. Further, it may solve the problem of misrepresentation in online dating.

On a general note, the present study is another example of how analyzing people's interactions in a virtually connected world allows for a deeper understanding of processes and behaviors much further than within the scope of computer-mediated communication.

## References

- Abrahams, M. (1994). Perceiving Flirtatious Communication: An Exploration of the Perceptual Dimensions Underlying Judgments of Flirtatiousness. *Journal of Sex Research*, 31(4), 283-292.
- Alexander, R.D. and Noolan, K.M. (1979). Concealment of ovulation, parental care, and human social evolution. In N.A. Chagnon and E.G. Irons (Eds.), *Evolutionary biology and human social behavior: An anthropological perspective*. Duxbury Press.
- Baltagi, B. (2008). *Econometric Analysis of Panel Data*. 4th edition, Wiley.
- Berger, C.R. (1988). Planning, affect, and social action generation. In L. Donohew, H. E. Sypher, and E. T. Higgins (Eds.), *Communication, social cognition, and affect*, Hillsdale, NJ: Lawrence Erlbaum.
- Berger, C.R. and Bell, R.A. (1988). Plans and the initiation of social relationships. *Human Communication Research*, 15(2), 217-235.
- Berscheid, E., Graziano, W. and Monson, T. (1976). Outcome Dependency: Attention, Attribution, and Attraction. *Journal of Personality and Social Psychology*, 34(5), 978-989.
- Brehm, S.S. (1992). *Intimate relationships*. New York: McGraw-Hill.
- Burgoon, J.K., Buller, D.B. and Woodall, W.G. (1996). *Nonverbal communication: The unspoken dialogue*. 2nd Edition. McGraw-Hill: New York.
- Buss, D.M. (1985). Human mate selection. *American Scientist*, 73, 47-51.
- Buss, D.M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, 12(1), 1-39.
- Buss, D.M. and Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology*, 50(3), 559-570.
- Buston, P.M. and Emlen, S.T. (2003). Cognitive processes underlying human mate choice: The relationship between self-perception and mate preference in Western society. In *Proceedings of the National Academy of Sciences*, 100(15), 8805-8810.
- Clark, R.D. and Hatfield, E. (1989). Gender differences in receptivity to sexual offers. *Journal of Psychology and Human Sexuality*, 2(1), 39-55.
- Clark, C.L., Shaver, P.R. and Abrahams, M.F. (1999). Strategic Behaviors in Romantic Relationship Initiation. *Personality and Social Psychology Bulletin*, 25(6), 709-722.
- Coyne, J.C. (1976). Toward an interactional description of depression. *Psychiatry: Journal for the Study of Interpersonal Processes*, 39(1), 28-40.
- Feingold, A. (1991). Sex differences in the effects of similarity and physical attractiveness on opposite-sex attraction. *Basic and Applied Social Psychology*, 12(3), 357-367.

- Fiore, A., Taylor, L., Mendelsohn, G.A. and Hearst, M. (2008). Assessing Attractiveness in Online Dating Profiles. In *Proceedings of the CHI 2008*, Florence, Italy.
- Frisby, B., Dillow, M., Shelbie, G. and Nordlund, J. (2010). Flirtatious Communication: An Experimental Examination of Perceptions of Social-Sexual Communication Motivated by Evolutionary Forces. *Sex Roles*, Online First, 18.08.2010.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.
- Green, S.K. and Sandos, P. (1983). Perceptions of male and female initiators of relationships. *Sex Roles*, 9(8), 849-852.
- Greer, A.E. and Buss, D.M. (1994). Tactics for promoting sexual encounters. *Journal of Sex Research*, 31(3), 185-201.
- Hancock, J., Toma, C. and Ellison, N. (2007). The Truth about Lying in Online Dating Profiles. In *Proceedings of the CHI 2007*, San Jose, CA, USA.
- Hitsch, G.J., Hortascu, A. and Ariely, D. (2006). What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating. MIT Sloan Working Paper 4603-06.
- Huffaker, D. (2010). Dimensions of Leadership and Social Influence in Online Communities. *Human Communication Research*, 36(4), 593-617.
- Joyce, E. and Kraut, R. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3), 723-747.
- Kelley, K. and Rolker-Dolinsky, B. (1987). The psychosexology of female initiation and dominance. In D. Perlman and S. Duck (Eds.), *Intimate relationships: Development, dynamics, and deterioration*, Newbury Park: Sage Publication.
- Kowner, R. (1995). The effect of physical attractiveness comparison on choice of partners. *Journal of Social Psychology*, 135(2), 153-165.
- Krebs, J.R., and Dawkins, M.S. (1984). Animal signals: Mind-reading and manipulation. In Krebs, J. R. and Davies, N. B. Sunderland (Eds.), *Behavioural Ecology: An Evolutionary Approach*. MA: Sinauer, 380-402.
- Lee, L., Loewenstein, G., Ariely, D., Hong, J. and Young, J. (2008). If I'm Not Hot, Are You Hot or Not? : Physical-Attractiveness Evaluations and Dating Preferences as a Function of One's Own Attractiveness. *Psychological Science*, 19(7), 669-677.
- Levinger, G. (1983). Development and change. In H. H. Kelley et al. (Eds.), *Close Relationships*, New York: W.H. Freeman and Company.
- Lightman, E.J., McCarthy, P.M., Dufty, D.F. and McNamara, D.S. (2007). Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. In D. S. McNamara and G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*, Hillsdale, NJ: Erlbaum.
- Little, A.C., Burt, D.M., Penton-Voak, I.S. and Perrett, D.I. (2001). Self-perceived attractiveness influenced human female preferences for sexual dimorphism and symmetry in male faces. In *Proceedings of the Royal Society B: Biological Sciences*, 268, 39-44.
- Nisbet, I.C.T. (1973). Courtship feeding, egg-size and breeding success in common terns. *Nature*, 241 (1), 41-42.
- Owen, J.E., Yarbrough, E.J., Vaga, A. and Tucker, D.C. (2003). Investigation of the effects of gender and preparation on quality of communication in Internet support groups. *Computers in Human Behavior*, 19(3), 259-275.
- Pennebaker, J.W., Booth, R.J. and Francis, M.E. (2006). *Linguistic inquiry and word count: LIWC*. Erlbaum, Austin, TX, USA.
- Regan, P.C. and Berscheid, E. (1997). Gender differences in characteristics desired in a potential sexual and marriage partner. *Journal of Psychology and Human Sexuality*, 9(1), 25-37.
- Reis, H.T. and Shaver, P. (1988). Intimacy as an interpersonal process. In S. Duck (Ed.), *Handbook of personal relationships*, Chichester, England: Wiley.
- Rellini, A.H. and Meston, C.M. (2007). Sexual desire and linguistic analysis: A comparison of sexually-abused and non-abused women. *Archives of Sexual Behavior*, 36(1), 67-77.
- Riessman, C. (2002). Doing justice: Positioning the interpreter in narrative work. In W. Patterson (Ed.), *Strategic Narrative: New Perspectives on the Power of Personal and Cultural Storytelling*. Oxford UK and Lanham MA: Lexington Books.

- Rude, S., Gortner, E.M. and Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121-1133.
- Stirman, S.W. and Pennebaker, J.W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4), 517-522.
- Tausczik, Y. and Pennebaker, J. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24-54.
- Thiessen, D.D. and Gregg, B. (1980). Human assortative mating and genetic equilibrium. *Ethnology and Sociobiology*, 1(11), 1-40.
- Todd, P.M., Penke, L., Fasolo, B. and Lenton, A.P. (2007). Different cognitive processes underlie human mate choices and mate preferences. In *Proceedings of the National Academy of Sciences*, 104, 15011-15015.
- Trivers, R.L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man*. Aldine.
- van Grove, J. (2010). Online dating is bigger than porn. <http://mashable.com/2010/03/24/online-dating-infographic>.
- Walster, E., Aronson, V. and Abrahams, D. (1966). Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology*, 4(5), 508-516.