

**Universität Bielefeld**  
Fakultät für Linguistik und Literaturwissenschaft

**Masterarbeit**  
im Studiengang Anglistik

zum Thema:

**The Truth about Mid-Life Singles in the USA:  
A Corpus-Based Analysis of Printed Personal Advertisements**

vorgelegt von

**Carolin Kirchhof, B.A.**

Erstgutachter: Dr. Thorsten Trippel  
Zweitgutachter: Dr. Stephan Gramley  
Bielefeld, im Juni 2010

# Table of Contents

1 Introduction.....	1
2 Theory.....	3
2.1 Central Concepts in Studying Personal Advertisements.....	4
2.1.1 Personal Ads.....	4
2.1.2 Sex vs. Gender.....	6
2.2 Personal Advertisements – Past and Potential.....	11
2.3 The Language of the Personals.....	16
2.3.1 The Surface Structure of Personal Ads.....	17
2.3.2 The Deep Structure of Personal Ads.....	18
2.4 Categorization.....	21
2.4.1 Classification of Personal Ads.....	21
2.4.2 Semantic Fields in the Personals.....	23
2.5 Stereotypes.....	31
2.5.1 Functions and Dysfunctions.....	31
2.5.2 Gender Stereotypes.....	33
2.6 Performance of Identity.....	39
2.7 From Speech to Writing.....	47
2.7.1 General Aspects.....	48
2.7.2 Modal Differences in Identity Negotiation.....	51
2.7.3 Stereotypes in the Personals.....	54
3 Empirical Part – Data Collection and Processing.....	58
3.1 Part-of-Speech Tagging.....	61
3.2 Terminology Extraction and Data Processing.....	65
3.3 Elicitation of Lexical Words.....	67
3.4 Results.....	68
4 Analysis and Discussion.....	71
5 References.....	74
6 Appendix.....	80

## Index of Illustrations

Illustration 1: Vplusfriends advertisement (Veltins 2006).....	13
Illustration 2: Betty Crocker advertisement (1971).....	35
Illustration 3: Distribution of Semantic fields across sexes.....	69

## Index of Examples

Example 1: Personal ad from <i>The Detroit Metro Times</i> .....	4
Example 2: Personal ad from <i>The Boston Phoenix</i> .....	13
Example 3: Personal ad from <i>The New England Courant</i> .....	14
Example 4: Personal ad from <i>The Hartford Advocate</i> .....	14
Example 5: Personal ad from <i>The Hartford Advocate</i> .....	18
Example 6: Personal ad from <i>The Hartford Advocate</i> .....	20
Example 7: Personal ad from the <i>NYT</i> , "Women seeking Men".....	27
Example 8: Personal ad from the <i>Detroit Metro Times</i> .....	30

## Index of Tables

Table 1: Personality trait adjective list (Hort et al. 1990).....	55
Table 2: Appearance trait adjective list (Hort et al. 1990).....	55
Table 3: Structure of Personal Advertisement.....	59
Table 4: Content word frequency in heterosexually directed personal ads.....	69

## **1 Introduction**

Classified advertisements appearing in newspapers are strongly restricted in their language not only with regard to number of words, but also to purpose, that is the message to be conveyed. Due to personal ads being a sub-group of classified advertisements, the definition of register given in Gramley (2001) can be transferred to them directly:

[T]hey clearly involve the goal of finding a partner (function); they are composed in an abbreviated type of language (style) using a code that includes conventionalized abbreviations and vocabulary from the area of personal relationships (field). (p. 202)

“[A]s personals in America reflect traditional values as well as gender differences and time-limited trends” (Parekh & Beresin, 2001, p. 227), they promise to be a resource for gender stereotypes with regard to cultural values and language. The ads are so limited with concern to word number, however, that linguistic gender stereotypes such as the usage of certain adjectives might not be as detectable as in spoken language. Personal advertisements promise to be a valuable resource in this regard because of the condensed performance of identity they provide. In addition to this, they are readily available for different genders, ethnicities, and locations; the specificity of the text type enables an efficient comparison. There is a lect inherent to US American personal ads. Its characteristics are clustering of initialisms, that is a group of initial letters forming abbreviations, strong reflection of stereotypes, brevity, unanimous structure and content as well as syntax and grammar. Content words in the advertisements will provide information on the topic of gender stereotypes since they carry the most referential meaning possible (cf. Bruthiaux, 1994, p. 26). Hence, this study will aim at the sociocultural characteristics in the self- and other-descriptions in a corpus of US American personal advertisements. By analyzing the semantically loaded words in a collection of more than five hundred classified personal ads taken from various US newspapers, these will be exploited for gender stereotypes in the context of desires with regard

to dating. The results are expected to reveal that sociocultural stereotypes are not significantly dominant in the advertisements. Instead, an elicitation of characteristics used by advertisers will demonstrate if and how reality complies with common gender stereotypes. Moreover, a gender-neutral lexical inventory is likely to be found in the data.

A theoretical section will introduce terminological aspects of this thesis as well as the employed conceptualizations of gender and gender stereotypes. The analysis of these in personal advertisements will be conducted on a corpus stored in a Tamino database. This means that, in contrast to relational databases (e.g. SQL, Microsoft Excel), the data will be accessible in tree-structured XML files. This provides for a more linguistically formal yet flexible annotation. The most frequently used lexical items will be grouped according to a selection of semantic fields related to those used by Gottburgsen (1995). These will encompass, among others, those of character traits, socioeconomic status, and outer appearance. After a description of the collection and processing of the corpus, the word frequencies will be assessed across genders. Following this, a comparison between the attested lexicons of heterosexual males and females will be conducted. The resulting preferences will be discussed in the course of this paper with regard to gender stereotypes such as 'women look for financial security' and 'men focus on looks'. Finally, a contrastive picture of the reality of personal advertisements will be drawn.

## **2 Theory**

In preparation for the corpus analysis, the theoretical background and research history of advertising aiming at personal partnerships will be reviewed. In this section, the terminological understandings of 'personal', 'sex', and 'gender' will be decided on. A brief historical overview of partner search will facilitate a better understanding of the material which is to be analyzed in the empirical study. Furthermore, the linguistic aspects of partner search advertisements with regard to their surface structure, that is textual design, and their deep structure, that is lexical makeup, will be discussed; a basis for the data processing and its evaluation will have been formed.

A short introduction of categorization as a concept for social and linguistic classification will be followed by a subsection on semantic fields. Based on a study of German partnership advertisements (Gottburgsen, 1995), lexical fields to be searched for in the present corpus will be formed. These shall later help to identify stereotype concentrations in the data. In order to decide on which stereotypes might occur in the corpus, several aspects of gender stereotypes will be examined. Moreover, after a general introduction to stereotyping, the display of gender stereotypes in advertising will be demonstrated. This will give further suggestions on what to expect in the corpus.

Identity performance is mandatory in this type of partner search and will, hence, be introduced here. Following this, an attempt at transferring linguistic gender stereotypes onto the written medium will be made. Sociolinguistic variationist research on speech, for example by Labov (1973) and Trudgill (1974), will be taken as a foundation. However, due to the specific register of personal advertisements (see section 2.3), no significant variation is expected in the data. A discussion of this will close the theoretical part, concluding with a statement of hypotheses on the data.

### ***2.1 Central Concepts in Studying Personal Advertisements***

### 2.1.1 Personal Ads

By definition, *personals* [emphasis added] are advertisements, and people are the commodities being advertised to be sold in a market where basic principles of supply-and-demand exist. (Parekh & Beresin, 2001, p. 226)

Nearly every newspaper, daily or weekly, has a section where one can place an inquiry for a potentially romantic relationship. The label for newspaper advertisements with the goal of finding a partner is not unified, but it is not very conflicting either. When skimming primary and secondary sources, one frequently encounters the terms 'matrimonial', 'personal', and 'personal ad' (e.g. Parekh & Beresin, 2001; *New York Times*), each of which will be discussed briefly in the following. Finally, a decision about the term to be used in this paper will be made.

In *The Vocabulary of World English*, Gramley (2001) speaks of *matrimonials* as those classified advertisements that are aimed at finding a partner (p. 202). He lists examples that express the search for a relationship; these do not necessarily express a wish for marriage. In the early stages of the research for the study presented in this paper, the terminology of *matrimonial* had been adopted because it seemed representative of the data. This is also the reason why the database contains the element name <Matrimonial> for the text items. After further research however, it became clear that *matrimonial* is mostly used for advertisements aiming at marriage (e.g. Nair, 1992). These are placed particularly often by Indian parents looking for husbands for their daughters. Example 1 is an announcement for a prospective wedding:

G GIRLS NEED LUV  
SBF, 31, **marriage-minded** [emphasis added], N/S,  
romantic, seeks man, 25-58, likes children, having fun, for  
friendship, maybe more.

*Example 1: Personal ad from Detroit Metro Times.*

As Divakaruni (2000) puts it in her overview of more than a quarter of a century of Indian matrimonials, “[m]arriage is a serious pragmatic commitment [...] involving significant financial transactions and requiring the blessing of parents and grandparents” (p. 3). Since only 11 out of 507 items (2.2%) in the present corpus contain the word *marriage* (e.g. Example 1, “Women seeking Men”), and none mention *wedding*, keeping the terminology of *matrimonial* would have been overgeneralization. Hence, another term had to be found.

As one of the major sources for the items in the corpus is the online version of the *New York Times* (2007), taking over their category seems plausible. The dating section of the newspaper is titled “Personals”, correlating with the Merriam-Webster (2010) explanation of the noun *personal*, that is “a short newspaper paragraph relating to the activities of a person or a group or to personal matters”. This definition is broad enough to encompass entries aimed at dating as well as those aimed at marriage. In order to avoid monotony in the text, *ad* or *advertisement* have been used to form compounds, respectively. Accordingly, *personal(s)*, *personal ad(s)*, *personal advertisement(s)*, *advertisement(s)*, or *ad(s)*, if not otherwise indicated, will from now on be adopted in this paper to describe the data.

Previous publications in the area of linguistics have also used this terminology (e.g. Cheshire, 1998; Parekh & Beresin, 2001). *Personal ads* is equally frequent in popular literature on the topic, for example in *Personal Ads: A career woman's approach* (Hale, 1995), and *Classified Love: A Guide to Understanding Personal Ads* (Marshall, 1994). The 'how-to' literature on personal ads is manifold, giving advice for finding a partner in a layman or even dubious fashion. This is at least the case when looking at from an academic point of view. While it may be entertaining at times, such literature is not very insightful as far as linguistics or sociocultural preferences of the authors in search of partners are concerned.



### 2.1.2 Sex vs. Gender

The usage of the terms *gender* and *sex* in this paper will be discussed and distinguished with the help of popular as well as subject-related encyclopedias (Merriam-Webster, 1994; Merriam-Webster, 2010; Ember & Ember, 2004). Formerly, the word *sex* would refer to a biological category, namely 'male' vs. 'female'. *Gender* would describe a grammatical one, that is 'feminine' vs. 'masculine' vs. 'neuter' (cf. e.g., Merriam-Webster, 1994, p. 417). These two terms, however, are increasingly often used interchangeably today, across both categories. *Sex* might evoke thoughts relating to its intimate connotation, which causes people to avoid this term. When one wants to know whether an infant is a boy or a girl, instead of saying, "What sex is the baby?", they might inquire on its gender as to avoid discomfort. In such a scenario, the biological, more specific domain would be left for the sake of consolation. Thus, the semantics of *sex* become more restricted while those of *gender* expand. The latter term is now combined quite frequently into lexical items like *gendered language*, *transgender*, or even *baby gender predictor* (Bounty, 2010). As can be seen in these three compounds already, *gender* has become far more than a grammatical descriptor. Nowadays, not only within academia, it refers to a person's performance of masculinity or femininity, that is of sexuality (see e.g., Segal, 2004). While biological sex markers are (usually) fixed, gender develops according to experience and preferences. Just as its referential terminology is, gender is a fluid concept and can change over time.

The blending semiotics between the grammatical, biological, and performance categories of the two terms over the last fifteen years is also reflected in the Merriam-Webster lexicons (1994; 2010). The encyclopedia (Merriam-Webster, 1994) has been specifically chosen because of the detailed explanations expected from it. The online dictionary (Merriam-Webster, 2010) on the other hand should supply briefer, more specific definitions because it is designed to supply more immediate information than a print dictionary. The growth in size and in number of definitions between 1994 and 2010 alone signifies the semantic expansion of *sex* and

*gender* over time. Within a span of ten years, each definition has gained at least two further distinctions over time, and the number of words has increased by 50%.

*Webster's new encyclopedic dictionary* (Merriam-Webster, 1994) already contains a rather extensive entry for *sex*. As its primary meaning, "either of two divisions of organisms distinguished respectively as male and female" (p. 939) is given. Furthermore, a closely related definition stands as a secondary explanation, namely the sum of all characteristics allowing for the prior distinction. Finally, sexual intercourse gives a circular definition to conclude the entry. One and a half decades later, the online dictionary entry for *sex* (Merriam-Webster, 2010) has been modified as well as it has been extended by two definitions. The primary and secondary explanations in Merriam-Webster (1994, p.939) have grown in word number, possibly due to their prior similarity. Furthermore, a semantic expansion of *sex* is notable: "[S]exually motivated phenomena or behavior" (p. XYZ) is added as a secondary definition on the level of 'sexual intercourse'. This suggests the acknowledgment of the above-mentioned performance category the term might spread to. The addition of 'genitalia' as a separate definition recognizes the semantic expansion from the biophysical *sphere*.

A similar phenomenon can be found in the widening scope of *gender* in the Merriam-Webster works of reference. Merriam-Webster (1994) defines *sex* and *gender* as synonyms in its primary definition (p. 417). In secondary position, *gender* functions as a grammatical distinction for nouns, pronouns or modifying word forms like adjectives. This agrees with the above-mentioned distinction of masculine, feminine, and neuter. What is more, the *Merriam-Webster Online Dictionary* (Merriam-Webster, 2010) supplies the user with a far more extensive entry for *gender*. Here, while the synonymy with *sex* has been moved to secondary priority, the explanation of being "a subclass within a grammatical class" (Merriam-Webster, 2010, "gender") has been promoted to first position. This definition is far more expansive and exemplifying than Merriam-Webster (1994), consisting of three sub-entries. Moreover, a remark on the arbitrariness of the attribution of said

gender categories to words, or rather objects, is made. Finally, in parallel to the entry on *sex*, “the behavioral, cultural, or psychological traits typically associated with one sex” (Merriam-Webster, 2010, “sex”) stands on one level with the synonymous definition. This explanation reflects the acceptance of the performance aspect of *gender* into common usage. Again, the growth in size of the definition over the years is undeniable with regard to the word number and to the number of different meanings.

At length, as far as dictionary definitions are concerned, the current popular understandings of the two terms in the context of people categorization are as follows: *Sex* is biological and dyadic, namely male or female whereas *gender* refers to the sociocultural attributes of and the identity performance by a person. Adopted from the grammatical terminology, *gender* is either masculine or feminine or a mixture of both ('neutral').

Several articles in the *Encyclopedia of sex and gender* (Ember & Ember, 2004) deal with the dual nature of the two terms from the point of view of the humanities. While Best (2004), for instance, decides to use *sex* and *gender* synonymously, Segal (2004) adds *sexuality* to the group. She makes sexual orientation a part of the construct. The latter trichotomy is plausible not only in its taxonomic organization but it also combines the common usage with the original meanings of the words. Segal (2004) explains that “[s]ex’ is taken to refer primarily to biological characteristics” (p. 3), that is to female or male physiology, which is inherent and stable (except for invasive surgery). This conforms to Merriam-Webster’s (1994; 2010) definitions above. “Gender’ is taken to refer to a culturally based complex of norms, values, and behaviors that a particular culture assigns to one biological sex or another” (Segal, 2004, p. 3). Accordingly, gender in Segal’s conceptualization is the sociocultural counterpart to sex. While every human being is born either male or female, “except for a few rarely occurring genetic or hormonal anomalies”, as Segal (2004, p. 4) correctly states, the way they are supposed to or do behave is shaped by their cultural and social environment. And, like the environment, this enactment

of gender is liable to change. This aligns with the definitions in Merriam-Webster (2010, v.s.) as well as with the usage advice by *Askoxford* (Oxford Word and Language Services, 2010, “gender”):

The words *gender* and *sex* both have the sense ‘the state of being male or female’, but they are typically used in slightly different ways: *sex* tends to refer to biological differences, while *gender* tends to refer to cultural or social ones.

Accordingly, all references to and analyses of gender language in this paper will refer to the self- and other-description of men or women. However, the categorization will not be split into masculine or feminine. Already Bem (1974) asserted that “[r]egardless of biological sex, an individual's gender role may be masculine, feminine, or some combination of both” (in Edwards & Hamilton, 2004, p. 494; cf. Laqueur, 1990). The latter then could be *androgynous*, that is it may have strong influences from both sides. Or, the gender role might be *undifferentiated*, namely not have many dominant characteristics from either category. This grouping of identities into gender-roles that one can choose and perform, allows for a more finely tuned analysis of behavior, especially when considering language.

What is particular about Segal's (2004) triad, though, is the addition of *sexuality*. It “refer[s] to the ways in which individuals structure their sexual and gender performances, and the partners toward whom they direct their behavior and emotional attachments” (Segal, 2004, p.3). Thus, *sexuality* in this context, as well as for this thesis, can be regarded as the sexual orientation of a person and the way it is performed. Finally, sexuality also is the way people perform, gender is what they perform. This performance is therefore, in parallel to the sexual partner preference, characterizable as either heterosexual, homosexual, or bisexual (cf. Segal, 2004, p. 5). Accordingly, the personal ads will be analyzed with regard to masculine and/or feminine identity. A special focus will be laid on language attributes and the stereotypes connected to these.

These divisions of sexuality, or sexual orientation, are largely paralleled by the categories of the personals in the present corpus. The sections “Men seeking Women” and “Women seeking Men” provide data of a heterosexual population and “Men seeking Men”; “Women seeking Women” supply the items of homosexual individuals. Bisexual orientation will be disregarded here because it is specified too scarcely in the advertisements; the abbreviation *Bi* occurred four times in the whole corpus, all used by women. Other forms of *bisexual* were not found at all. People also might not regard such indication necessary because of the category they advertise in. Apart from this, the fact that a person advertises themselves in one of the four previously listed categories indicates their current partner preferences.

## **2.2 Personal Advertisements – Past and Potential**

Today, nearly every newspaper or magazine offers personal ads, including free and non-free, daily, weekly or sporadic publications. In addition to this, as is the case with the data in the corpus used for this paper, many publishers put their classifieds with all subsections on the Internet for public access. Looking for a partner via advertisements nowadays is fairly easy due to the availability of far-reaching media. According to Steinfirst and Moran (1989), “[t]he growing demand since “the 1980s created an explosion of the personals” (as cited in Parekh & Beresin, 2001, p. 224). This popularity is still palpable today (e.g. Groom & Pennebaker, 2005).

The history of public partner search, however, can be traced back as far as the 1800s (Parekh & Beresin, 2001, p. 224). As has been described in the novel *Waltz into Darkness* by Cornell Woolrich (1947, as William Irish), ordering a bride via mail from Britain was quite popular at least in parts of the United States, namely in the New Orleans of the 1880s. Parekh and Beresin (2001) give a condensed overview of the diachronic development of personal advertisements. They also mention “verbal advertising” (p. 224) as a common means for finding a suitable partner, albeit for marriage, in the early days of the U.S. Not only in the form of peddling on the market place, this form of advertising was

[i]nitially used by religious groups and minority populations, match-making eventually diverged from the path leading to today’s personals. Soon after the advent of newspapers [*sic*], written personals appeared. (Parekh & Beresin, 2001, p. 224)

The first newspaper was printed in Germany in 1605 (Kilman, 2005, ¶15). More than a century later, one of the first personal advertisements in North America was placed by Benjamin Franklin on April 23, 1722 (see Example 3) in the *New England Courant*, of which his brother was the

editor. Used for arranging extramarital affairs at first, the personals developed into a common medium for finding a partner.

The demand in many US Americans for finding a partner, new or as an affair, grew from the 1980s onward (cf. Parekh & Beresin, 2001, p. 224). The reasons for this were and still are, among others, increasing divorce-rates, less going out due to 'home-entertainment' and longer working hours, as well as frequent change of residence. This demand is still mirrored in the immense numbers of personals in print and online, as well as in the growing partner-mediation industry. Furthermore, the emancipation of women had an effect on the number of personals placed. As the average age in the present corpus reflects, the average age of singles advertising has risen over time (cf. Parekh & Beresin, 2001, p. 224). In the data analyzed in this paper, 64.1% of the advertisers are 40-59 years old, possibly including divorcées. According to Parekh and Beresin (2001), the growth of newspaper usage for finding a partner might also be caused by the strong influence of appearance in meeting places like bars, etc. Singles might try to avoid by presenting themselves more favorably in writing (cf. Parekh & Beresin, 2001, p. 224).

The authors also mention fear of infection with sexually transmitted diseases (STD) as a disadvantage for meeting partners in pubs or clubs (Parekh & Beresin, 2001, p. 224). In contrast to the insecurity in a casual acquaintance, in a personal one is already supplied with the essential information on a person. However, apart from some occurrences in the section "Men seeking Men" (*HIV negative*, Example 2), the avoidance of mentioning STD is dominant in the present corpus. This fact does not support the supposed advantage of advertising over going out that is suggested by Parekh and Beresin (2001). Their idea that "columns of personal advertisements offer a quick and decent overview of compatible mates" (p.225), in contrast, is an advantage compared to having to approach people first to learn more about them. Accordingly, the personal ad technique is "a big time-saver" (Parekh & Beresin, 2001, p. 225) which can be comfortably placed from home via the telephone or the Internet.

HEALER, ARTIST...  
 thinker? GM, youthful 50, handsome 5'11",  
 160lbs, **HIV** - [emphasis added], laid-back  
 yet accomplished, possessed of deep soul,  
 intellect, SOH, into cycling, film, dining,  
 travel. Seeking 45-55 -year-old any race,  
 background, for romance and more.

*Example 2: Personal ad from The Boston Phoenix.*

Today, personal ads are publicized in many varieties besides the usual short texts in newspapers. One can find them online, for instance, in text form or as even as videos (e.g. Sun, 2006). Furthermore, they can be in the form of extensive profiles on dating platforms, including pictures and explicit listings of hobbies, taste in music, etc. Examples for these websites are *eHarmony* (<http://www.eharmony.com/>) and *A1 Indian Matrimonials* (<http://a1im.com/>). The latter platform is an Indian as well as Muslim platform intended to find marriage partners specifically, preferably in the United States of America. In Germany, the brewery *Veltins* even let people advertise for partners or friends on the bottles of their V+ product series in 2006 (Illustration 1), including a picture, profile, and anonymized contact data :



*Illustration 1: Vplusfriends advertisement (Veltins 2006).*

What is more, a single advertisement is not only available through just one source anymore. This is also the case with the data used for the present study. Most newspapers and magazines have a freely accessible online edition in which the personals are also published. Apart from the fact that the e-procuring of the personal ads is convenient, they are a rich



source for humanistic study because of their content. Parekh and Beresin (2001) describe their content as follows: “Every aspect of personals gives information – from the language used, to the paper in which they appear, to the societal and cultural contexts driving them” (p. 230). Personal advertisements depict the current use of language and other cultural tendencies, enabling diachronic as well as synchronic analyses of these aspects. This is one of the reasons why the material has been chosen for this study. It lends itself in an excellent way to research gender stereotypes and language. As Parekh and Beresin (2001) postulate,

[p]ersonal advertisements are powerful windows into understanding individuals, societal trends, and cultural values. 'Personals' in the United States and elsewhere offer a unique opportunity to understand societal changes and cross-cultural issues. (p. 1)

Synchronic studies can allude to current issues in a certain culture while diachronic analyses can depict societal change. When comparing, for instance, Benjamin Franklin's rather jokingly placement in 1722 (Example 3) with an advertisement from 2007 (Example 4), changes in the written language can be noticed.

Any young Gentlewoman (Virgin or Widow) that is minded to dispose of her self in Marriage to a well-accomplish'd young Widower, and has five or six hundred pounds to secture [sic!] to him by Deed of Gift, she may repair to the Sign of the Glass-Lanthorn in Steeple-Square, to find all the encouragement she can reasonably desire.

*Example 3: Personal ad from the New England Courant (Beauman, 2010).*

A REAL GENTLEMAN SBM, 36, 5'11", 210lbs, Taurus, light brown eyes, mustache, never married, no children, medium build, neat, well-groomed, attractive, truck driver by trade, enjoys casinos, day trips, quiet drives, camping, the water, quiet moments, concerts, and occasional nights out. ISO compatible, caring lady.

*Example 4: Personal ad from The Hartford Advocate.*

Apart from the fact that the intention of marriage in the early 18<sup>th</sup> century seems stronger than two hundred years later, the undertone of the man helping the woman, or “poor widow” is not present anymore in 2007. This gives prospect to detecting gender stereotypes, or rather a condensed performance of gender, in a corpus analysis. The lack of initialisms in the older advertisement as well as the more syntactical expressions used are striking. Personals are indeed a very specific kind of textual matter and, thus, supply fertile grounds for linguistic analysis.

### **2.3 The Language of the Personals**

Biber (1988) distinguishes between *genre* and *text type* (p. 170). The classification according to function instead of form makes a piece of writing belong to a genre. Items with different composition share the same purpose. Personal advertisements are very fixed with regard to design and lexical makeup, which can be seen in a high regularity of overlapping attributes. Accordingly, these ads, which aim at finding a partner, belong to one text type. This again is a sub-class of advertisements in general, especially classified ads. Every text type not only has unique characteristics of its register, it is also distinct in format.

As has been mentioned briefly in the introduction, the language used in personal advertisements is a certain type of register. The functional tenor, or particular purpose of personal ads is very concrete and of provident character, namely finding a partner in the “real world” (cf. Gramley, 2001, p. 187). Gramley (2001) also sorts them within the group of directives, which “are concerned with concrete future activity. Central to such texts are imperative forms with an imperative effect [...]” (p. 188). Pleadingly aimed at the partner-seeking readers of the classified section, the ascription of directive seems fit. As a further specification, the text form *announcement* can be ascribed to the personals, since a bold statement of being a single ‘in search of’ is involved. (cf. Gramley, 2001, p. 189).

The language of personal advertisements has been discussed by several linguists with different foci. Bruthiaux (1994) and Cheshire (1998), for instance, are concerned with lexical and grammatical variations in personal ads of the Western world. Nair (1992) and Divakaruni (2000), in contrast, focus on idiosyncrasies of Indian matrimonials in English-speaking venues across the world, researching cultural differences. Furthermore, an ever growing community of researchers is looking into the linguistics of online dating. Among the study foci are the language of emails as well as that of profiles on dating websites (see, e.g., Groom & Pennebaker, 2005; Fiore & Donath, 2005). Apart from being characterized

by the brevity of personals, and their restricted register (cf. Gramley, 2001, p. 202), Nair (1992) notes the following about their potential:

This apparently peripheral genre not only displays a characteristic content relating to social stereotypes (...), it also shows distinctive features of lexis, syntax, and discourse organization (...). A study of this subvariety of written language is of theoretical interest both because it has considerable human and social implications and because in form and style it is quite amenable to strictly linguistic analysis. (p. 232)

It promises to be a very rewarding venture to exploit what personal ads can tell about stereotypes with regard to sociocultural as well as to linguistic aspects. Since they are already grouped according to sexual direction, gender stereotypes especially will be elicitable. Hence, this paper is concerned with the language of printed personal ads and what it may convey about gender stereotypes.

The characteristic discourse organization of personal advertisements will be described in the following section, as well as some of its lexical and syntactic features. The majority of ads is structured similarly and they also share a common dictionary of fixed expressions and abbreviations. The latter are nearly always initialisms.

### **2.3.1 The Surface Structure of Personal Ads**

The layout of the personals section is usually very simple and clear. It may be several pages long or just occupy a small section of the classifieds pages. The ads are grouped into categories of sexual orientation, namely two heterosexual and two homosexual ones. However, one or the other might be left out depending on the readership of the newspaper. Each category forms one or more vertical columns topped by a header. The items are separated by thin black lines and ordered by the liking of the editors. Example 5 from the "Personal Ads" section of *The Hartford*

*Advocate* (2007), placed in the category “Male seeking Female”, is a very typical instance of American personals:

**A GOOD MAN 4 U SM, 60,**  
creative, caring, compassionate, likes home  
cooking, classic rock, writing. WLTM similar  
lady to share talks, casual dates, friendship  
and fun. Possible LTR. Box 18728.

*Example 5: Personal ad from The Hartford  
Advocate.*

The title is bold and supplies essential information about the advertiser. A list of personal details is succeeded by what kind of lady the author *WLTM*, that is, “Would Like To Meet” (see section 2.3.2). A long-term relationship is the goal of this personal; interested women may contact a post box number which is registered with the publishing newspaper.

No matter which category, the greater part of US American personals in printed and electronic papers have the same core elements (see Example 5, v.s.): A boldly printed headline with an eye-catching word or phrase which usually includes a statement of the advertiser being single. In the same-sex sections, the sexual orientation is often added. This is followed by the ethnic heritage, gender, and age of the author in abbreviated style, and their most-defining psychological and physical attributes. After a list of preferred leisure time activities and other personal description, the desired features of the partner being searched for are indicated. These also include gender, age, and often physical or intellectual properties of the (potential) addressee. Frequently, the desire for a lasting relationship is also expressed. The previously listed content may vary in order of appearance but its essentials are omnipresent. An ad usually closes with a postbox number or an email address to contact the composer.

### 2.3.2 The Deep Structure of Personal Ads

Going deeper into the structure of personal ads, the deciphering begins. In example 5, the initialism *SM* stands for “Single Male” and is part of the most frequently used ones in American newspapers. *WLTM* means “Would Like

To Meet". Other recurring shortenings are *S* for 'Single', *W* for 'White', *B* for 'Black', *F* for 'Female', and *LTR* for 'Long-Term Relationship', etc. (cf. section 3.4). *S* occurs externally from abbreviations in 87 of the 507 ads in the present corpus already while 520 initialisms include an *S* overall. The majority of *S* stands for single, other cases include 'smoker' or parts of geographical shortenings like *SF* for 'San Francisco'. The word form of abbreviation, especially its subclass of initialism, however, gives fertile ground for further research as a core feature of personal ad language.

It is striking that personals in printed American newspapers largely are enumerations of attributes regarding the people involved. In contrast to many of their German counterparts, they do not have grammatical features such as a sentence structure. Moreover, the only verbs written out in full are scarce and mainly restricted to instances of 'seeking', 'searching', 'looking for', and others with similar meaning. The most common replacements of these verbs are *WLTM* and *ISO*, the latter standing for 'In Search Of'. Similarly noticeable is the fact that conjunctions like *and* or *or* are often either omitted or replaced by symbols such as *&* or */*.

Another major characteristic of American personals is the indication of religious belief. Especially in the *New York Times*, being Jewish (*J*) is often stated in the self-description and required in the potential partner. Similarly, *P* is included quite frequently in both parts of the placements, demonstrating the relevance of Protestant beliefs. While religion is often an important criterion, ethnicity is seldom a requirement, though. However, the statement of skin color or ethnicity is another obligatory aspect in the personals, including *J* for 'Japanese' and *H* for 'Hispanic'.

While the core aspects, namely an introduction to the content part in the form of capital clusters and statements on character, are mostly obligatory, there do not seem to be any restrictions on creativity. Some people include lines of poems in their placements, others quote from their favorite movie. The most important part seems to be the bold printed heading to catch the eye of the reader in the flood of usually more than fifty advertisements per

page. In contrast to the content, this heading has to be immediately noteworthy. After having fulfilled this requirement, the personality of the advertiser is more important than style of the advertisement. Generally, a brief personal description followed by expectations from a partner are essential. However, the style in which this information is presented varies. Example 6 (“Female seeking Male”) demonstrates how both head and body are very appealingly designed:

ME, MY ABC'S  
Artist, Berkshires, Cocky, Delightful, Eclectic, Funny,  
Gardener, Hippie, Imagination, Jamaican Sunsets, Kitty  
Cats, Lover, Matisse, New Mexico, Opinionated,  
Painter, Quirky, Rainbows, Scorpio, Theater, Upbeat,  
Vulnerable, White Wine, X rated, Young At Heart, even  
in my 60's, Zest For Life - and you? BOX 1276841  
*Example 6: Personal ad from The Hartford Advocate.*

While looking catchy, this ad is unstructured with regard to content. By using nearly only capitalized adjectives and nouns, it gives the impression of a to-do list rather than of something heart-related. Moreover, the advertiser does not include any desired attributes of the person she is searching for. The expected train of thought in composing a personal, though, would be "What do I want from a partner", followed by "What would she or he like in me?". This is suggested by the majority of the advertisements in the present corpus. Accordingly, an advertisement should include the answers to these questions. It seems that, in addition to an inviting outward appearance of the surface structure, the appealing presentation of the content is crucial to the success of a placement. However, since the authors of the analyzed advertisements are anonymous, one can only guess at the success because of the frequency of this pattern in the corpus.

## **2.4 Categorization**

Categorization according to preselected markers is a widely discussed topic in variationist sociolinguistics. Benwell and Stokoe (2006), among others, mention the problem that most researchers first construct a framework and then attempt to fit speakers that have been chosen specifically for their potential to fulfill this purpose into it. Such a procedure causes circularity in so far as it only proves anew that, for example, ethnics perform ethnicity by talking about ethnic concerns (Benwell & Stokoe, 2006, p. 56). By this means, no new aspects can be discovered. Furthermore, most analysts do not go far beyond their predefined categorization and thus minimize the scope of relevance. With this procedure also comes the problem of perspective, since "performativity studies [...] rely heavily on *analysts'* rather than *participants'* categories" (Benwell & Stokoe, 2006, p. 57). A speaker might not choose the identity they have been labeled with by a linguist. More importantly, they might not perform it intentionally or, for that matter, fittingly. Equally, they might not concentrate on behaving according to an actively self-selected identity while speaking, but rather express it unconsciously. As has been found by Benwell and Stokoe (2006), neither women nor men would necessarily label their language as gendered. Rather, linguists choose groups already categorized according to biological markers hoping to find corresponding linguistic markers.

### **2.4.1 Classification of Personal Ads**

For the purpose of eliciting established gender stereotypes from the present corpus of personal ads, however, certain choices have to be made. Since the advertising population is anonymous, no follow-up questions can be asked. The design of a personal, however, is a highly conscious process in which the authors attempt to express their ideas on what is important about their identity. The conveyed image may be exaggerated but at the same time it is highly expressive. The most important aspect here, however, is the aim of the ads, namely finding a partner of the same or



opposite sex. This is signified by the category the text is put in. Thus, we expect people to perform largely according to their sexual identity (see 2.5.4) indicated by the category in which they place their announcement. Accordingly, the categories already given in the newspapers have been adopted.

Following the definition of *gender* given in section 2.5.1, the performance of sexual identity can have feminine and masculine characteristics. In order to avoid possible gender conflicts, only those ads of heterosexual advertisers will be included here. Consequently, a top-down, semiotic approach in the fashion of Coates (1986; cf. Stubbs, 1983) has been adopted. Another reason for this choice of method is that the biological sex of the authors is already made clear through the categorization provided by the newspapers while their gender is not. This categorization by sexual orientation is inescapable when examining personal ads. It will be tested in a separate paper, though, whether the language used in the given set of personals does in fact conform to the four predefined categories, namely “Men seeking Men”, “Men seeking Women”, “Women seeking Women”, and “Women seeking Men”. One could speculate, for example, that some homosexual men might use more femininely stereotyped language than heterosexual men do, etc. This, however, would have to be investigated using a bottom-up, informational approach. Within the data, disregarding the greater context, the individual words and structures would be used to construct an interpretation of their constellation (cf. Stubbs, 1983). In such a case, all vocabulary occurring in the personals would be sorted intuitively, disregarding who used it. Later, one can check who leads in which assorted lexical field, regardless of their gender or ethnicity. Afterwards, new categories could be deduced from the results, which would exclude the influence of stereotypes. Furthermore, one might check the assorted fields for gender preferences by revealing their ethnographic information. Since both of the approaches mentioned above would exceed the focus of this thesis, the data analysis will be restricted to the heterosexual categories already present in the data.

A further form of categorization, albeit a subjective one, is the grouping of lexical words into semantic fields. Allwood (2006) explains the purpose of lexeme classification according to shared sentential aspects as follows:

The basic idea in studying a semantic field is to collect words that are similar in meaning and then try to analyze the relationship between them or try to compare a semantic field in one language with a semantic field in another language. (p. 46)

Among the best known examples for this methodology is the comparison of color terms across languages and cultures (e.g. Berlin & Kay, 1969 ; Wierzbicka, 1999). Lexical fields are conceptual domains and “[...] consist of ensembles of near-synonymous lexical items” (Geeraerts, 2002, p. 7). The items within a semantic field can be compared with regard to number and variation, for instance, allowing to reach conclusions about its users.

In order to allow for an analysis of gendered preferences with regard to vocabulary use in the personals, the words of the present corpus have been sorted into semantic fields. In the following section, a categorization model of lexemes occurring in personal advertisements will be developed based on the one proposed by Gottburgsen (1995) for German personals. Not only does she focus on gender performance in the personals, she also presents a detailed classification of lexical items in this text type. As has been discussed above, semantic fields allow for cross-lingual comparison and thus have the potential to be applied cross-culturally as well.

### **2.4.2 Semantic Fields in the Personals**

The study of gender stereotypes in the personals requires an analysis of content words. According to Lutzeier (2006), those POS contributing most to lexical fields “in terms of size and structures are found for nouns, verbs, and adjectives” (p. 80). Hence, as has been done by Gottburgsen (1995), the collected data will be filtered by these word classes.

In her study *On the Presentation of Gender through Language – Doing Gender in Personal Columns*, Gottburgsen (1995) analyzes personal advertisements from the German broadsheet *Frankfurter Allgemeine Zeitung (FAZ)*. Based on intuition and peer-review, a set of semantic fields that are prone to be used most frequently by both, men seeking women and women seeking men was assembled (Gottburgsen 1995, pp. 272-273). The daily newspaper publishes a strictly heterosexually oriented section in its weekend edition under the heading of “Ehewünsche/Partnerschaften” (F.A.Z. Electronic Media GmbH, 2010). The most basic placement currently costs about €100, making the clientèle a rather exclusive one. Gottburgsen (1995) proposes that the communicative act of putting a personal ad in the paper, and the goal expressed by this, leads the author to perform their gender and, thus, to stick to stereotypes (p. 269).

In her study of the FAZ, Gottburgsen (1995) looks for instances of the following semantic fields: *appearance* (Aussehen), *social behavior* (Sozialverhalten), *emotionality* (Emotionalität), *charisma* (Ausstrahlung), *eroticism* (Erotik), *gender specificity* (Geschlechtsspezifität), *adequacy* (Adäquatheit), *education* (Bildung), *vocation* (Beruf), *financial assets* (Vermögen), *talent* (Begabung), *rationality* (Rationalität), and *mental disposition* (Psychische Disposition) (cf. Gottburgsen, 1995, pp. 272-273).

Since a bottom-up design of the present study would have taken to many resources (v.s.), the top-down approach used by Gottburgsen has been partially adopted. The development of a comparable set of semantic fields, even with the help of *WordNet* (cf. Miller, 1995), would have been equally subjective. Thus, the concepts introduced by Gottburgsen (1995) will be taken over with some modification in order to be searched in the present data. The newly combined lexical fields will be presented with explanations and exemplary lexical items. The results of this paper and those concluded by Gottburgsen (1995) will not be compared here. A widening of the scope to cross-cultural relations would, however, be an interesting topic for further research.

In the following, the lexical fields used in the study of US American personal advertisements presented in this paper will be explained. The semantic field of *appearance* consists of lexical items concerning a person's physical attributes (cf. Gottburgsen, 1995, p. 272), and how these are perceived by others. Words in this field can be either concrete (e.g. *blonde* (hair color), *5'3"* (height), etc.) or subjective (*attractive*, *sporty*, etc.). The descriptions can be either in detail or rather commonplace, but in general are employed for the purpose of conveying a positive picture. The lexical items will be elicited from the corpus without considering collocations or compounds. This is the case because combinations of, for example, instances of *long blond hair* as a descriptor item is far less frequent than *blond hair*, *long hair*, or *hair*, respectively. Hence, words such as *eyes* or *body* will equally be sorted into this semantic field because a collocation with an appearance-describing lexeme is obligatory (e.g. *blue eyes*, *slim body*). Especially regarding instances of *eyes* or *hair*, the probability of the single lexemes is far higher than their color specifiers and thus are more expressive with regard to the frequency of the field of *appearance*. Following the same principle, modifiers such as *good-* and *well-* (as in *good-looking* or *well-built*) will also be categorized into this semantic field.

Words belonging within the field of *social behavior* describe how a person behaves around others, or how they affect the community. Examples for this group are *reliable* or *humorous*. Lexemes that can be attributed to *charisma* express what an effect a person has on others, for instance whether they are charming or nice. While Gottburgsen (1995, pp. 272-273) separates these two semantic fields, they are closely related because both are concerned with how a person is perceived during human interaction in general. Hence, the combined field of *social behavior/charisma* will be used in the present study.

A field closely related to that of *social behavior/charisma* is formed with lexemes belonging to the category of more intimate human interaction, namely *emotionality/eroticism/relationship*. Again, Gottburgsen (1995, pp.

272-273) separates these three, with *emotionality* including words pertaining to feelings and actions caused by them. *Eroticism* then encompasses items regarding a person's behavior during body contact. For both fields, Gottburgsen (1995) gives the example of *zärtlich* (tender), while attributing *liebevoll* (gentle) to the first, and *tolerant* to the second (pp. 272-273). Not only does this categorization seem rather arbitrary, but it is also flexible. It appears that lexical items belonging to the semantic field of *eroticism* can easily be placed in that of *emotionality*, and vice versa. Therefore, only one field encompassing *emotionality* as well as *eroticism* as distinguished by Gottburgsen (2005, pp. 272-273) will be used in the following. Moreover, the lexemes in this sentential grouping are all somewhat concerned with a romantic *relationship* (cf. Gottburgsen, 1995, p. 273) between the two prospective partners searched for with the advertisement. Accordingly, those items fitting into this larger semantic field will also be included here, forming *emotionality/eroticism/relationship*. Specific to the latter lexical field is the compound *long-term relationship*, or its abbreviated form *LTR* (see Example 6). As is the procedure with collocations in *appearance*, occurrences of *long-* or *term* will be treated as part of *LTR* at all times. Instances of *time* or *times* will be treated accordingly since they are bound to occur with modifying adjectives such as *good* or *long*. Therefore, they all belong within the semantic field of *emotionality/eroticism/relationship*.

While already having been grouped into sections of “Women seeking Men”, etc., the vast majority of advertisers mention their own sex and that of the person they are looking for again. In the body of the advertisement, this seems to be an obligatory convention. In the self- and other-descriptions, this is often expressed via the subjects, that is agents, of the abbreviated-style sentences. While the majority of classified advertisements are not grammatical, that is are lacking syntax, they all have an agent of some kind. This can either be in the form of an introductory initialism, such as SBF (single black female), or explicitly and multiple, as in Example 7. These abbreviatory lexemes, which seem as

mandatory in US American personal ads as in the *FAZ* sampling (Gottburgsen, 1995), will be sorted into the semantic field of *gender specificity* as defined by Gottburgsen (1995, p. 273). As has been discussed in section 2.1.2, *gender* in this paper normally refers to the performance of sexual identity. In the classification of denotative lexemes, however, the terminology used by Gottburgsen (1995) will be adopted for its grammatical origins. *Gentlemen*, for instance, is grammatically masculine, *woman* is feminine, etc.

DO YOU HAVE A ZEST FOR LIFE  
 SWF [emphasis added], 51. Slim, upbeat, LI lady  
 enjoys theatre, museums, nature, spirituality and  
 more. Seeks happy SPM [emphasis added], N/S,  
 tall, sincere gentleman who is reflective and likes to  
 hug. For LTR. BOX 18392

*Example 7: Personal ad from the NYT, "Women seeking Men".*

Further instances of this semantic field are *lady*, *guy*, *female*, etc. Moreover, all capital letters within initialisms will be grouped here (s.v.). A slightly more detailed, though not encompassing discussion of abbreviations in the present corpus is given in the two final paragraphs of this section.

Gottburgsen (1995) also uses the category of *adequacy*, encompassing information on whether a person is “entsprechend, angemessen oder übereinstimmend” (p. 272). Items grouped here relate to anything that both, the author and the addressee, hope to have in common with a potential partner. As Fiore and Donath (2005) report,

[p]sychologists have found that actual and perceived similarity between potential romantic partners in demographics, attitudes, values, and attractiveness correlate positively with attraction, and, later, relationship satisfaction. (p. 4)

*Appearance* (including attractiveness), which also factors into compatibility, is a large enough semantic field to make up an autonomous

category. Hence, the other aspects regarding similarity of the potential partners will be grouped in an additional, encompassing field. Gottburgsen (1995) only gives two literal examples of lexemes belonging to her field of *adequacy*, namely *adäquat* (*adequate*) and *passend* (*fit* or *fitting*) (p. 272). Nonetheless, this sentential grouping should include a wider variety of words expressing adequateness that are more subjective. The extended version of the semantic field used in the present study, that is *adequacy/compatibility*, will cover all words concerning the compatibility of the potential partners in many areas. For instance, in addition to the rather abstract examples given by Gottburgsen (1995, p. 272), other variables regarding factors important to compatibility in general will be grouped here as well. Examples of these are hobbies (e.g. *computer games*, *visiting art galleries*) and plans for life (e.g. *live out by the sea*). What is more, occurring parts of collocations such as *same things* (*same* and *things*) will be treated as belonging within this field. *Things* will supposedly occur more often than instances of *good things* or *similar things*, respectively. It is therefore more telling about the frequency of the semantic field of *adequacy/compatibility*.

The lexical field *education* will contain terminology relating to general and academic education as well as to upbringing. As with *appearance*, Gottburgsen (1995, p. 273) suggests that both, vague and specific lexemes belong in this field. Accordingly, concrete items like *academic* and *graduate of XY University* will be treated as equals of, for example, *sophisticated* or *traditional*.

Partly overlapping with *education* are the semantic fields of *vocation* and *financial assets*. In her study of German personal ads, Gottburgsen (1995, p. 273) separates the two subcategories. There, *vocation* contains lexemes such as job designations as *lawyer* or *cook*, and circumscribing terms, for instance *successful*. *Financial assets* include *well-endowed* or (*financially*) *independent*. These two areas are closely connected as occupational success leads to higher wage. What is more, the mentioning of one might make the other obsolete. Accordingly, the semantic field of

*vocation/financial assets* will be combined for this study of US American personals.

Gottburgsen (1995, p. 273) also lists *talent* (Begabung) as another semantic field to be expected in personal ads. Her definition, “,Begabung': bezeichnet die natürlich Anlage, angeborene Fähigkeit zu bestimmten Leistungen (Kernwörter *kreativ, musisch*)” (p. 272), seems plausible. Accordingly, lexemes describing non-mental and non-social predispositions of a person including, for example, *artistic* or *musical* belong to this field. In ambiguous cases, words will be grouped in both, *talent* and *adequacy/compatibility*, since the latter also includes hobbies.

Moreover, the two categories of *rationality* and *mental disposition*, which are treated separately in Gottburgsen (1995, p. 273), have been merged to form the last semantic field for the analysis of the present data. *Rationality* encompasses lexemes relating to claims about a person's intellectual and rational abilities (e.g. *smart, intelligent, rational*). *Mental disposition* provides information on the “psychischen Kompetenzen, die als Voraussetzung z.B. für den Berufserfolg oder für das Meistern des Lebens überhaupt aufzufassen sind [...]” (Gottburgsen, 1995, p. 273). The latter includes *independent, zest for life, ambitious*, etc. Due to the fact that both semantic fields contain words relating to mental matters and sociability, a unification into *rationality/mental disposition* has been decided on.

Finally, the aspect of ethnicity (or race/phenotype/religion/etc.) remains to be dealt with. It is not mentioned by Gottburgsen (1995), possibly because the *FAZ* readership is a rather homogeneous group of upper middle class citizens. Lexemes describing this category in the present corpus will be grouped in the field of *adequacy* because their mentioning implies their importance to the author. In US American personals it is very common, or even mandatory, to place an introductory abbreviation at the beginning of the text body, either in the title or within the first line. In example 8, SWM introduces the advertiser a 'single white male' (N/D=non-drinker). His potential mate is a SF, a single female.



**AN EYEBROW RAISER?**

SWM, 57, N/D, smoker, computers and crafts,  
seeks SF, 45-60, for friendship and possible LTR.

*Example 8: Personal ad from the Detroit Metro Times.*

The statements of singleness as well as of gender, while being obligatory, are redundant since they are already obvious through the fact that the ad is actually placed and in the corresponding category of “Male seeking Female”. Aiming more at ethnic compatibility is the skin color of the composer, that is the W in SWM. There is a great variety of ethnic groupings in both self- and other-description. These include DWPM (dark white, i.e. Hispanic Protestant male) or SW/AF (single white or Asian female). They reach from skin color (B, W, DW, etc.) to religion (J for Jewish, P for Protestant, C for Christian, etc.). However, some letters can have double meaning. Depending on the position they take in an initialism they change in meaning. An example for this is *J*, which can be Jewish, Japanese, or something else, depending on sociocultural consensus or the guidelines of the newspaper.

The exact construction and deciphering of these often very complex abbreviations will not be dealt with in the empirical part of this paper, though, since they are so conventionalized and ubiquitous that they do not allow for variation across sexes. The only aspect to be mentioned here is that the same-sex advertisements contain the letter 'G' at some position in the abbreviation, while heterosexually directed personals seldom include sexual orientation. However, there could be an introductory “POSSLQs,” which means 'persons of opposite sex sharing living quarters'” (Parekh & Beresin, 2001, p. 226). This convention is as superfluous as the stating of one's own or the potential partner's sex and sexuality because of the category in which the ad is placed. However, it is just as obligatory.

## 2.5 Stereotypes

[...] Person A chose[s] these items as self-descriptive: attractive [sic], dependent, emotional, gentle, kind, talkative. Person B chose[s] these items: active [sic], ambitious, determined, inventive, self-confident, serious. [...] Is it easier to imagine one of these individuals as a man? Which one? As a woman? Which one? Is it easier to visualize Person A as a woman and Person B as a man? If so, your views demonstrate the influence of gender stereotypes – beliefs about how men and women differ in their psychological make-up. (Best, 2004, p. 11)

In this section, the construction of stereotypes by society as well as how they are performed by individuals will be discussed. The inescapable presence of markedness and cross-cultural pragmatic differences will also be addressed. Finally, the focus will be laid on gender stereotypes, preparing the analytic basis for the study of US American personal ads.

### 2.5.1 Functions and Dysfunctions

McGarty, Yzerbyt, and Spears (2002), for example, propose that “(a) stereotypes are aids to explanation, (b) stereotypes are energy-saving devices, and (c) stereotypes are shared group beliefs” (p. 2). Taking these assumptions together suggests that overgeneralization makes life easier by ascribing attributes to others. However, they differ depending on the sociocultural environment and on the “angles of telling” (Singh, 1999, p. 98). For one, the in-group/out-group phenomenon is relevant in this context. Furthermore, cross-cultural variation, for instance regarding politeness, makes stereotyping much more practicable. This cognitive process is often referred to as a separation between *us* and *them*, that is the relationship between 'stereotyper' and 'stereotypee'. We perceive others by differences to ourselves and label them accordingly. Singh (1999) posits that this happens through “negative labeling” (p. 98). She describes processes of stereotyping, specifically focusing on ethnicity. However, considering McGarthy's (2002) functions of labeling for simplification (s.v.),

the observations made by Singh (1999) can be generalized nearly without exception. Men frequently think about women that these talk too much. In contrast to that, men are often termed to be too pragmatic by the other sex. It seems rather unlikely that a man would describe a woman as talking 'very explicitly and detailed'. Nor might a woman possibly identify a man as 'being so nicely precise'. At least this would be an assumption going against common stereotypes.

What is more, making statements marked by stereotypes “suggests that the intended audience [...] shares the same beliefs and attitudes” (Singh, 1999, p. 99). This again relates to the sociocultural circumstances of applying stereotypes. A label put on someone has to be recognized by the interlocutor, a “perceptual link” has to be established (cf. Singh, 1999, p. 99). The phrase 'you talk like a woman' directed at a man might mean that he is either scarce in words or overly talkative, depending on the cultural context. Furthermore, putting a stereotype label on someone can evoke either laughter or aggression exactly because of cultural practice. Allport (1990) gives reason to assume, however, that stereotypes function as “labels of primary potency” (as cited in Singh, 1999, p. 100) and are an expression of power. They suggest a dominance of the speaker over the subject (e.g. example on 'you talk like a woman', v.s.).

A notable reaction to this display of superiority by negative labeling is the “attempt[...] to 'take power back' by reclaiming such terminology” (Singh, 1999, p. 101). An example for such a reconquer is the usage of the term *gay* after its transgression of primarily meaning 'happy' to being a denotation for 'homosexual'. The change in meaning of *gay* is described in Harper (2010) more or less chronologically. Entries with meanings that are too closely related are omitted in the following summary. While in 1178 *gay* meant 'full of joy or mirth', in 1637 its mere utterance suggested immorality. At least as far as the United States of America are concerned, the connotation of male homosexuality was first noted in the late 19<sup>th</sup> century. Then, *gay cat*, for instance, was used for migration workers who spent long whiles among solely male comrades. In 1920 already, “The 'Dictionary of

American Slang' reports that *gay* [...] (adj.) was used by homosexuals, among themselves [...]” (Harper, 2010, “gay”). Finally, in 1971 *gay* was noted as a noun describing a usually male homosexual. Harper notes the recapture of the term by the male homosexual community in 1920. At this point, the negative descriptor formerly expressing power over a minority (e.g. *gey cat*, 1893) was turned into a positive label of solidarity. Today, *gay pride* has become a symbol of power within the LGBT (lesbian, gay, bisexual, and transgender) culture, completing the semantic inversion by (re-)establishing cultural and linguistic practice.

A similar reversion is reported by Singh (1999, p. 102) on the reclamation of the label *nigger* for African-American men as displayed in the movie *Jackie Brown* by Quentin Tarantino (1997). Apparently, the director Spike Lee, himself an African-American, criticized Tarantino for demanding the solidary use of *nigger*. As a member of the 'white' majority group Tarantino might imply even more negative connotations, where these were not intended. Examples like this “show that it is difficult to reclaim certain labels totally as positive markers of [...] identity” (Singh, 1999, p. 102). It seems that the dyad of power and solidarity is inescapable and impassable in the context of negative labeling, and stereotyping. This is especially so because solidarity as well as power are expressed through “linguistic practices which mark them as distinctive” (Singh, 1999, p. 103; see also Gramley & Pätzold, 2004, pp. 195 ff., for a more detailed discussion of *power* and *solidarity*). The binary opposition is omnipresent. Andersen (1988) explains that “[i]n attempting to reclaim abusive terminology [...], [minorities] are rejecting the labels and norms imposed by the majority and 'taking power back'” (as cited in Singh, 1999, p. 110).

### 2.5.2 Gender Stereotypes

The dual systems of masculine-feminine and homosexual-heterosexual are dominant in North America and Western Europe. Other cultures, for instance that of the Plains Indians, have a broader system (cf. Segal, 2004). In Thailand one can adopt a second official gender in addition to the

congenital one, namely *ladyboy*, or *kathoey* (Loxton, 2007). In spring 2010, the state of New South Wales, Australia, acknowledged the first person to be neither man nor woman (Gibson, 2010). Apparently, gender is a social construct used to categorize people. A construct like the Western dichotomy “constrains and directs understandings of sexual behavior, sexualized behavior, and their association with nonsexual aspects of social and cultural life” (Segal, 2004, p. 4). Gender stereotypes are channeled through a sociocultural construct and constantly adopted, applied, and rejected. As a result, identity is enacted according to one's self-negotiated gender role. This, in turn, is “a factor that influences the interpretation of messages” (Edwards & Hamilton, 2004, p. 1; cf. Scotton, 1980) because the addressee of a statement will use their cultural knowledge for interpretation.

Upbringing and education familiarize us with certain ideas of what women and men are supposed to be like (Scotton, 1980; cf. Krueger, Hall, Villano, & Jones, 2008). This means that growing up teaches how to react to self- and group-representation as well as how to perform. Mothers put their daughters in pink and purple clothes because that is 'what girls look like'. Little boys wear blue. Apart from experiencing such gender performance at home, at school, and in other social settings, part of what is regarded as typically male or female is suggested by the media. Movies or soap operas already present an exaggerated version of life. In the sitcom *Home Improvement* (Finestra 1991-1999) of the 1990s, for example, manliness is defined through working on the mechanics of cars and handiwork, though on a highly ironic level. Some commercials as well as print advertising, however, display gender roles on an even more overstating level. Often, the distributors of products hint at sex roles as expected by society. Examples for this are the brands *Betty Crocker* (see Marks, 2007) and *Aunt Jemima* (see Manring, 1998); both US American labels which sell baking mixes and cooking helpers as well as ready-made food. Illustration 2 depicts a *Betty Crocker* advertisement from 1971 (Typepad.com, 2008):



The text above the cake mix box reads, “When he compliments you on your homemade frosting, why bother him with unnecessary details?”. Not only does the story depicted suggest that a woman should please her husband. She also should pretend that this is “easy as pie”. While current advertising is not as stereotypical anymore, the shiny face of a happy housewife is still present on the cover of every product *Betty Crocker* sells (Marks, 2007).

It would seem that [at least] the print media continues to be in thrall to sex difference discourse which perpetuates conventional assumptions about men and women and which treats departure from gendered scripts as deviance. (Day, Gough, & McFadden, 2004, p. 334)

While advertisements for household products often personify the stereotypical woman as a good cook and housewife, men are often displayed in advertising as specifically expecting women to make food for them. Hentges, Bartsch and Meier (2007) conducted a study on the way and proportions of men and women in television commercials. For that purpose, they reviewed studies on the roles of male and female stereotypes as represented in such commercials. First of all, Hentges et al. (2007) found out that “[a]lthough there is some indication that gender stereotyping is declining (Bartsch et al. 2000), [it] is still prevalent” (p. 55).

Apparently, stereotyping has an impact on the attitudes and behavior of adults. In fact, Hentges et al. (2007) report that “both cognitive-developmental and social learning theories propose television as a source of information on gender roles” (p. 56). Studies also revealed that women in commercials are, in general, more attractive than their male counterparts. In addition to this, they over-represented in this way (cf. Hentges et al., 2007, p. 56). More importantly though, in reference to stereotype acquisition, Hentges et al. (2007) attest that sex role suggestions are also prominent in television commercials aimed at children. These might even contain more stereotyping than commercials with adult targeting. Indeed, “girls are more often found in domestic settings [...], and [...] voice-overs [the omniscient narrators of TV commercials] are predominantly male” (Hentges et al., 2007, p. 56). This might suggest a sort of dichotomy of subordination and dominance, which would be quite a stretch. The constant reinforcement of stereotypes through the media does, after all, take part in children's “process of figuring out what it means to be male and female” (Hentges et al., 2007, p. 57). It might either reinforce or contradict the real life experience, resulting in an unconscious or conscious process of identity formation. Hentges et al. (2007) could confirm some of the suggestions in a range of television commercials directed at different age groups. They found that 64% of all the product representatives were female while 54% were male (p. 58). Moreover, due to the high percentage of male characters in general, there were also more men in product authority roles, that is decision makers on purchases. This overrepresentation of certain stereotypes is suggestive of women seeing men as naturally being in charge. Such a projection is especially noticeable in commercials aimed at school children:

[In these there are] nine times as many male authorities as female authorities compared with only 1.58 times as many for commercials from adult programs. Advertisers use significantly more males to 'sell' products to school-aged children. (Hentges et al., 2007)

The mediated dominance of male authority figures in television commercials could be suggestive to children that this is the societal norm. Also, the same might be true for the dominant display of women in a domestic setting. However, this can only be guessed at. There is still a need for studies in how far gender stereotypes suggested on television, especially in the commercials impact children, or adults. Also, it has not been proven yet whether the assumptions made by those wanting to manipulate the audience with these stereotypes are correct in the first place (cf. Hentges et al., 2007, p. 61). Nonetheless, it is a fact that the classification of the domestic space as female and the financial and the authoritative spaces as male is outdated.

However, since stereotypes are helpful for quick classification, they probably are what comes to mind first when making new acquaintances (cf. McGarty et al., 2002). The following statement by Singh (1999) on ethnic labels can be transferred to a more general concept of stereotyping:

[T]he ideologies of ethnic majority groups become established as norms, and [...] everything that does not conform is represented, and perceived, as different and peripheral. This also holds true in the context of representations and perceptions of ethnicity. [...] This is typically achieved in discourse by explicitly creating an opposition between *us* and *them* [emphasis in the original] and making use of negative labelling and stereotyping. (p. 98)

Every individual has certain idiosyncratic ideas of what masculinity or femininity means to them just as of what they consider 'snobby' or 'cool'. We categorize others according to these ideas when stereotyping by applying positive or negative labels to *them*. And this system is what children acquire from the beginning. They go through a stage of overgeneralization in which they might name every quadruped 'doggie' or every bird 'duck' (see e.g., Hopper, 1972; Crystal, 2003). Similarly, every person holding a briefcase in their hand will be just like 'mommy' or 'daddy' if this is what the parents look like frequently. This 'recognition of signs'



goes on later in life. For instance, a person in a black cloak and a white collar might be classified as a priest on the first look, a boy with damaged pants and a Mohawk as a punk, etc. On the second look, however, this first impression might be contradicted or falsified. And this can be transferred directly to gender stereotyping. For example, a man with an elaborate hairdo, tight jeans, and a pink shirt would have been stereotyped as a homosexual some years ago. Actually, he possibly still would be by some. The new lifestyle of 'metro-sexuality', however, gives men the space to care about their looks while still being a fully acknowledged heterosexual (cf. Castelo-Branco, Huezco, & Lagarda, 2008). Furthermore, women have been wearing pants for several decades now. In the 1920s, however, this would have been frowned upon, because these garment were considered a male pejorative during these times. Nobody then would have considered men wearing skirts either, for that matter, which can be seen occasionally nowadays (cf. Kaneko, 2010).

All in all, the male and female spaces keep blending further. Not only in the domains of fashion, work, and domestic and social life, the male-female dichotomy of stereotypes seems to be valid. Nevertheless, it keeps being applied. After an outline of the processes of linguistic identity performance (2.6), the possible blurring of stereotype borders in this area will be discussed. Since personal advertisements are restricted with regard to space and content, they will show a high degree of identity performance (2.7).

## 2.6 Performance of Identity

[I]n early 2005, an Internet search on 'identity' reveals a preoccupation with 'identity fraud', 'identity cards', and 'identity theft', all of which point to a common-sense use of the term as something that people own; a personal possession that can be authenticated or falsified. (Benwell and Stokoe, 2006, p. 17)

Singh (1999) also states that “[identity] is something we all have, and it is either part of mainstream norms or marked as distinct from those norms” (p. 94). She also says that language is an overt symbolization of identity. Its interpretation is always only one 'angle of telling' (cf. Singh, 1999, p. 93). Common issues of identity, especially regarding the linguistic performance of the latter, are discussed by Benwell and Stokoe (2006) in *Discourse and Identity*. As researchers of variationist sociolinguistics, they “theorise [...] identity in a similar way to social identity theory, as a pre-discursive construct that correlates with, or even causes particular behaviours” (p. 26). The variationist discipline of sociolinguistics deals with how messages are conveyed and perceived differently with, for example, alternation in grammar or the lexicon. This variation is not only restricted to a formal-vs.-informal dyad. For instance, 'Would you be so kind as to give my a second copy of the paper?' transmits different signals about the interlocutors' relationship than 'Gimme that!'. Similarly, prosody in speech can function as a social marker, for example by using a condescending tone towards an employee. A similar goal can be reached by increasing lexical and syntactical complexity in writing as is the case with legalese. Finally, non-verbal behavior such as gestures, posture, and mimics function as channels for attitude and identity (cf. Halliday, 1989, pp. 30-32). However, “[c]hange in language use does not mean immediate change in attitude” (Singh, 1999, p. 101). Factors influencing these linguistic variations are social markers like gender, age, class, education, cultural identification, etc. Trudgill, for example, conducted research on pronunciation patterns (e.g. Trudgill 1974), while Coates analyzed gender language (e.g. Coates,

1986). How theories about identity performance can be applied to the written medium will be discussed in 3.6.1.

One of the most prominent works on the language(s) of men and women is the *Tannen Model of Gender Communication*. Tannen (1990; 1994) postulates that the communicative behavior of the two sexes is crucially different. She states that women aim at solidarity during linguistic performance. Men, on the other hand, supposedly act in accordance with relations of power and dominance. Tannen (1990; 1994) induces from this that initial unidirectional goals in interaction are always maintained. Men always negotiate power while women always talk to reach or keep a solidary relationship; a cooperative-controlling continuum exists. This divergence in language codes should cause a near omnipresence of misunderstandings in cross-sex dialog. As Maltz & Borker (1982) put it, “[i]t is, in essence, cross-cultural communication” (as cited in Edwards & Hamilton, 2004, p. 491). What has been overlooked, or rather ignored, by Tannen (1990; 1994) is that the male-female dichotomy has long been outdated. Moreover, if men would always try to secure a dominant, controlling position for themselves more than just verbal cross-sex communication would encounter difficulties. Edwards and Hamilton (2004) tested and successively re-evaluated the Tannen Model of Gender Communication (Tannen, 1990; 1994). An important assumption for the reexamination was that the application of gender stereotypes would influence how messages are perceived. Sex only figured in to a minimal degree. As Bem (1974) defines, “regardless of biological sex, an individual's gender role may be masculine, feminine, or some combination of both” (as cited in Edwards & Hamilton, 2004, p. 494). This allowed for more variation and a more flexible ascription of linguistic variables. Tannen (1990; 1994), however, allowed no mixture between or within sex and gender in her model. Also in contrast to Tannen, whose “notions [...] were based on anecdotal evidence and analyses of small numbers of individuals” (Edwards & Hamilton, 2004, p. 1), 192 subjects were tested during their re-evaluation.

Commenting on Bakan (1966), Edwards and Hamilton (2004) presupposed that “masculine [...] qualities include self-assertion and control, whereas feminine [...] qualities include concern for others and emotional expressiveness” (p. 493). Their results showed that androgyny was implied when masculine as well as feminine qualities were present to a high degree in a subject. In contrast to this, “those with low levels of both traits [...] [were] labeled 'undifferentiated'” (Edwards & Hamilton, 2004, p. 494). This proves the point that there are more gender roles than just either wholly masculine or feminine. In fact, Edwards and Hamilton (2004) used parts of the Bem Sex Role Inventory (BSMI; Bem, 1974) for gender role assessment. “According to Bem (1987), a sex-typed individual is someone whose self-concept incorporates prevailing cultural definitions of masculinity and femininity” (Hoffman & Borders, 2001, p. 40). Hence, the inventory assembled by Bem (1974) is appealing to be used for studies on gender stereotypes. Bem's framework, while being widely criticized for being underdeveloped and misinterpreted (e.g. by Gilbert, 1985; Frable, 1989; Hoffman & Borders, 2001), does include the broader psychological division into feminine, masculine, androgynous, and undifferentiated. However, apart from possibly being outdated in 2010, the BSMI is supposedly rather to be used to diagnose androgyny than anything else (Frable, 1989). For this reason, as well as due to other criticism (v.s.), the BSMI itself will not be used for this study on personal advertisements. Nonetheless, all of the four psychological gender categories used by Bem (1974) show in the data elicited by Edwards and Hamilton (2004). In turn, it provides meager support for Tannen's (1990;1994) model. Finally, the researchers prove that indeed, men are neither from Mars not that women are from Venus (Edwards & Hamilton,2004). In fact, they promote an escape from the bipolarity of gender. Even so, the researchers also suspect that “the popular media seem to emphasize intrinsic differences between the sexes, which may lead individuals to believe that improved communication is impossible” (p. 503). This might cause people to be more reliant on stereotypes in studies they participate in (cf. 2.5.1).

As has been discussed above, Tannen (1990; 1994) assumes gender roles to be omnipresent, inescapable, and to be causal to speech and communication at all times. The general assumption that the perceived relationship between language and identity is (always) causal, however, has been criticized by Cameron (1997, as cited in Benwell & Stokoe, 2006, pp. 26-27), among others. Edwards and Hamilton (2004) also deem gender to be influential only. Linguistic choices are not constantly made consciously. The crux of this criticism is that identity is rather negotiated during performance than fixed; language is individually motivated (cf. Scotton, 1980). "As the different characters enter and affiliations change, so do the *us* and *them* groups" (Singh, 1999, p. 98). Such a development happens when, for example, a former student is now teaching at university. The student community then changes from *us* to *them*. The same reversal of relations happens when one travels to a foreign country and becomes the foreigner oneself. What is crucial to the expression of identity is the environment in which it takes place. While societal constructs are of great influence in any negotiation of identity, the situational factors are even greater. For instance, one's behavior at work is different from that around friends. Sometimes, gender or sexuality is actively denied in one place while being proudly enacted in another. In heteronormative societies like most Western cultures, this may be the case of a homosexual working in a homophobic environment. Equally, a heterosexual accompanying friends at a gay-pride event will try to fit in and thus align his behavior accordingly. Local identity changes at all times.

In whichever situation, identity is negotiated and performed in agreement with the "dynamics of each interaction" (Scotton, 1980, p. 359). As has been mentioned above, even in a situation where the hierarchy is predefined, violations of the norm can occur. Through alignment in communication, we change roles like actors in different plays. This character performance is also reflected in the language domain. As Scotton (1980) claims, "[m]aking linguistic choices [even] is one of the most 'visible' aspects of role-taking" (p. 359). In every conversation, the statuses of the

interlocutors are present. These may be displaced on a hierarchy that is defined by age or SES, for example employer – employee or salesperson-customer (cf. Scotton, 1980, p. 361). The interlocutors might also be on the same level in the hierarchical order, for instance student – student. In such predefined relationships, the situational identities are pre-established and hence are usually not challenged. When, however, two parties meet for the first time, in parallel to the status relationship, the situational identity has to be negotiated. Here, “linguistic choices [are] synonymous with symbolising status” (Scotton, 1980, p. 359). The interlocutors will use politeness strategies to converge and to communicate successfully. These strategies might involve hedging (e.g. Dixon & Foster, 1996) and other methods in the realm of Face Theory (Goffman, 1963). For instance, indirectness will be preferred to bold requests in order to ensure polite communication (see also Brown & Levinson, 1987). Moreover, the newly acquainted should follow Grice’s Maxims (Grice, 1975) to ensure a positive outcome. Whichever means is used or neglected, the “linguistic choices [speakers make] can be explained as individually-motivated negotiations of identity” (Scotton, 1980, p. 360). One signals attitude and individuality through behavior in a conversation. This is accomplished, for instance, via lexical and grammatical ‘performance’. Hence, the politeness strategies listed above do definitely play a large part in face-to-face communication. The form of communication treated in this paper, personal advertisements in newspapers, unfortunately does not allow for direct reactions to a proposition utterance. But, while the negotiation of identity can be described as procedural, the resulting performance is always present in language.

Nevertheless, parts of identity, be it sexual, ethnic or other, might be disguised by a speaker as a result of negotiation. It might even be faked and bear no resemblance to the actual identity, for example by someone who imitates a foreign accent. Following up on the theory of “identity-as-construction” as proposed by Coates (1986), Benwell and Stokoe (2006) agree that speakers have access to different identities which they can

alternate between or mix. For instance, if a devout Catholic priest would gleefully join the soccer-watching crowd at a pub he surely would change his language. Obviously, identities are performed differently in different social settings, not only with regard to non-verbal aspects. They are influenced by accommodation strategies and are realized through a change in verbal and non-verbal behavior. Accordingly, a label previously attached might cease to fit in a new situation, resulting in a correlation fallacy on the part of the researcher (Benwell & Stokoe, 2006, p. 26). A conflict between self- and other-ascription based on stereotypes occurs. This can also be said for the relationship between sex and gender as it is not fixed. Furthermore, "the fact that [...] identity can incorporate many different characteristics means that its definition is neither clear-cut nor uniform" (Singh, 1999, p. 97).

As has been discussed in section 2.4 of this paper, the specific aim and restricted register of the medium of personal advertisements promises an intense performance of identity. The individual categorization of those who advertise themselves in a newspaper suggests a choice of sexual identity by the performing individuals. The language chosen for this performance is, according to Scotton (1980), always either marked or unmarked within a normative framework (p. 360). She (1980) defines these choices as follows:

[T]he unmarked choice is that choice which the norms of society indicate represents the most expected choice for a particular status-holder in a particular role relationship in a particular situation [...]; it is the most expected because it is, in fact, the choice most often made.

In relation to what has been mentioned above about communicative situations, linguistic performance fitting sociocultural norms is applicable in clearly defined status relationships. At least when conforming to politeness constraints, utterances will likely be unmarked most of the time. It has to be noted, however, that this is not necessarily a conscious process (cf. Scotton, 1980; Benwell & Stokoe, 2006). Nevertheless, marked choices have a potential to be employed intentionally. This might even be expected

by societal norms (cf. Gumperz, 1976 in Scotton, 1980). The cause for marked statements might also lie in the speaker's intention to loosen up the situation, or to be critical toward the interlocutor. Yet, when the social hierarchy has to be negotiated first, so will the situational identity of the interlocutors. In such a “weakly define [sic!] role relationship” (Scotton 1980, p. 361), the utterances made are influenced by 'exploratory choice'. When marked, these can be assumed to be less interpreted as such, in contrast to occurrences in pre-defined situations. In general, however, all linguistic choices made will be evaluated by the addressee. This usually happens in accordance with the framework of the given sociocultural norms. Based on the linguistic performance, Scotton (1980) proposes that “members of the speech community will form a consensus in defining the speaker's negotiated identity” (p. 362). While an agreement like this may exist in language varieties such as AAVE or teen slang, this definition is more likely to be situation-dependent. At least, this can be said for said weakly defined role relationships as described by Scotton (1980). Furthermore, since identity is continuously negotiated depending on the interlocutors, a consensus of a whole speech community would rather imply the application of stereotypes than the recognition of and individual's identity. Scotton (1980) agrees with this phenomenon, stating that “[r]ole expectations will bias the interpretation of the marked choice so that it will be viewed as a dis-identification” (p. 361). As in the example of the priest in the pub, interpretation is relative to the communicative situation. Deviations from stereotypes can be understood differently depending on how well the interlocutors know each other. The application of stereotypes will seize diametrically to the increase of intimacy.

Scotton (1980) expands her categorization of weakly-defined role relationships further, putting linguistic choices in relation to the goals of the interlocutors. The interlocutors may have short- or long-term goals that can be different on either side. Their identity negotiation, and with it the linguistic performance, will be influenced by what the speakers aim at in a communicative situation. In order to reach certain goals, interlocutors will



follow the 'Gains Maxim' (Scotton, 1980, p. 363). This implies that one might ignore societal norms and make marked choices to be successful momentarily. An example for such behavior is the use of condescending or ridiculing language toward a person on the same status level so as to win a discussion. The marked choice of the speaker will signify temporary dominance while violating the norm. A situational suspension of the marked/unmarked distinction is used for higher aims. A stronger variant of the Gains Maxim would be a complete situational denial of identity. This has been exemplified with the negation of sexual identity in a hetero- or homo-normative environment, respectively, as well as with the priest in the pub. The higher aim for acceptance suppresses normative behavior. An alignment in identity performance can also be found in written communication. There, however, it is due to a variation in the availability of linguistic means.

## **2.7 From Speech to Writing**

Through most of the 20<sup>th</sup> century, and especially since the distinction of *langue* and *parole* by de Saussure (1916/1983), linguistic research has focused on spoken language. Scholars such as Bloomfield (1933) disregarded the written medium as “merely a way of recording language by means of visible marks” (p. 21). Later on, the study of texts starting in the 1950s, among others by Chomsky (1957), was rather of a structuralist nature. Only from the 1970s onwards, more deeper scientific research on the written medium has come into view (e.g. Halliday 1978). During this time, another evolution came with the functional approach to language, for instance by the Prague school (Vachek 1973) and by Trudgill (1974). Horowitz (1987) describes the advances of this shift in focus as follows:

Theoretical papers and volumes with this view show that language and texts [sic!] grow out of human needs to construct, negotiate, and interpret meaning for an audience and the personal intentions of a speaker or writer. (in Horowitz & Samuels 1987, p. 122)

Finally, the written modality has been granted a purpose of its own. In doing so, a continuum has been defined reaching from conversational speaking to academic writing (cf. Chafe & Danielewicz in Horowitz & Samuels, 1987). Nevertheless, most literature on sociolinguistic variation is still almost exclusively concerned with spoken language. However, several authors (e.g. Halliday, 1985; Biber 1988; Stubbs, 1980; Tannen, 1982) have discussed the relationship between speech and written language. This provides a basis for an adaptation of theories on spoken language. Unfortunately, nothing has been published yet on the transfer of gendered language onto the written medium of personal advertisements. This section will attempt to bridge this gap in order to make assumptions about the varieties to be expected in the present corpus. A summary of relevant literature will be given, noting the core aspects of written-language variation.

### **2.7.1 General Aspects**

Crystal (2005) gives a comprehensive overview of the similarities and differences between spoken and written language. For one, speech is temporarily restricted and dynamic, and thus transient. Moreover, speech is mostly used in face-to-face communication or when, apart from the speaker, at least one interlocutor is physically or otherwise present, such as on the phone. This may be a private conversation, a public political speech, or even a podcast. Accordingly, utterances are made with one or more specific addressees in mind. Writing, on the other hand, apart from letters or emails, often has a more widely spread readership. Books, articles, or web-blogs may be intended for a specific audience, but they are, at least in principle, accessible to everyone. This is due to the sustainability of written communication, which is contrasted by the lag between reception and production in speech; recordings do make an exception here. While speech is more spontaneous, writing is planned and editable. What is more, speaking does not allow for rephrasing, only for immediate self-correction. Editing is possible and done while writing. In speech, on the other hand, too much repair will have a negative effect on the communicative efficiency of an utterance. The written medium can take close to infinite attempts to be completed. It might not even be finished or made accessible to any addressee at all. The potential disadvantage of this lag between production and reception is the inability of the author to perceive the reaction of the addressee. Hence, “care needs to be taken to minimize the effects of vagueness and ambiguity” (Crystal 2005, p. 150). Equally, extralinguistic as well as co-expressive facial expressions and gestures are restricted to immediate interaction. The lack of these modalities forces the author to be more explicit because the text is self-contained. More planning, anticipation, and speculation on the effects of the utterance is involved than in face-to-face communication. As Chafe and Danielewicz (1987) found, “[a]s a result, written language, no matter what its purpose or subject matter, tends to have a more varied vocabulary than spoken” (in Horowitz & Samuels, 1987, p. 86). This fact supports even more the idea that identity

is constructed especially densely in personal advertisements. They are pieces of text with a high lexical density which are meticulously planned and thought through. But, as Chafe and Danielewicz (in Horowitz & Samuels, 1987) assume, “there is a discontinuity between what people have in mind and the language they use to express it” (p. 87). Nonetheless, the more carefully planned usage of words will come closer to what a person wants to express than spontaneous utterances. This has been discussed thoroughly by several linguists (e.g. Levelt, 1989; Gahl, Garnsey, Matzen, & Fisher, 2006, but will not be further considered in this paper. The mere number of possibilities to express a thought will have to suffice to prove this discontinuity.

Another difference between the two modalities is that in writing no situational references such as manual, spontaneous deixis can be applied. “[O]ther paralinguistic features, including aspiration, laughter, voice quality; timing, including simultaneous speech” (Stubbs 1980, p. 117), etc. are limited to the spoken medium. This lack in the written modality may be detrimental to comprehension (Horowitz in Horowitz & Samuels 1987, p. 127). Another aspect of writing is that topic changes have to be made chronologically if communication is supposed to be successful. Intonation and pauses organize information differently than punctuation does and can. Crystal (2005) also mentions the difference between the two channels within general syntactic structure. While speech employs concatenation, writing prefers embedding (cf. Halliday in Horowitz & Samuels 1987, p. 73). Structuring signals in texts like *therefore* or *however* may be helpful in texts but might seem overly formal in speech. Nevertheless, organizational means such as layout or graphics can also support communication, albeit on a different level. At the same time, their styles are in a continuum-like relation (Tannen, 1982). While writing supplies the most formal end, the least is formed by casual spoken communication (see also Gramley & Pätzold, 2004, chapter 1.4). Furthermore, certain expressions are idiosyncratic to either speech or written language, the latter more often than not being more formal and standardized.

Generally, functions that both media can supply may correspond indirectly or directly, or overlap (cf. Stubbs 1980, p. 117). Also, the feature persistence is possible in writing because of its situational restrictions. While speech is used in casual and unplanned discourse, writing almost always allows for a more structured discursive organization. Crystal (2005) gives sub-clauses as well as syntactic and lexical complexity as a characteristic of written communication. For example, a feeling of surprise is conveyed differently in both modalities. In speech, one might gasp and increase pitch and volume while continuing the conversation. In writing, on the other hand, a verbose explanation of the emotion has to be given, e.g. 'I am so surprised to hear you got married!'. This way, not only the temporal length of the utterance expands but also the spatial one.

What is more, writing is generally more suitable for factual interaction, be it via a coffee machine manual or an anthropology of Greek literature. It is supposed to be clear and concise. Speech is more casual and, at the same time, more final. Errors made can be corrected but not erased. A slip of the tongue can cause laughter or crying, respect or derision. Such consequences cannot be planned or avoided in spontaneous speech; speech writers may think differently about that, though. However, questions by the listener can be answered directly, feedback is not delayed or absent. On the contrary, "written language [...] is essentially one-directional" (Stubbs 1980, p. 117). This makes face-to-face communication potentially more effective for certain topics. On the long run, however, the sustainability of the written medium is greater (s.v.) than that of speech. The permanent presence of texts make better means for actually studying what has been proposed by the author. In the case of personal ads, for instance, the decision to reply allows more pondering than spontaneous contact at a cafe. Furthermore, the actual creation of a piece of writing is similar to a speaker-internal conversation. The editorial process allows for playing through different scenarios and for trying out different formulations and speculate on their effect on the reader. Then again, the permanence of text requires a more careful creation than spontaneous speech. Stubbs

(1980, p. 111) comments on the explicitness in writing with regard to its distinctive code as proposed by Bernstein (1971):

[R]estricted code has been defined consistently as implicit, particularistic and context-bound, because it relies on shared understandings between speakers and shared knowledge about contexts; whereas elaborated code is, conversely, explicit, universalistic and context-free, because it has to stand without such shared understandings.

The usage of a form of elaborated code thus seems mandatory for personal ads. While the context or aim of the placement, namely partner search, may be known to both author and addressee, the interlocutors themselves are not. No audience involvement is possible (cf. Chafe & Danielewicz in Horowitz & Samuels, 1987, p. 105). As in every piece of text, the writer has “to present content in a form which follows and supports the ideas expressed [...] the match between form and content should ensure efficient and effective understanding” (Horowitz in Horowitz & Samuels 1987, p. 127). The type-token ratio in speech is thus lower, the number of expressive words higher in written language due to the variety of lexical choice (see Chafe & Danielewicz in Horowitz & Samuels, 1987, p. 88). This also makes less common words and phrases more frequent or even unique to writing than in speaking. Abbreviations, for instance, are largely restricted to the written medium. Especially the subclass of initialisms are frequent in the present corpus of personal advertisements.

### **2.7.2 Modal Differences in Identity Negotiation**

As far as identity is concerned, it is directly recognizable in speech by the listener. The ethnicity, gender, age, etc. of the speaker are immediate. As has been discussed in the previous section, identities can be manipulated as well as alternated between. However, through means such as pitch, rhythm, accent, tone, etc., the speaker has a wide range of parameters to perform their identity. Or, facial expression and gestures are

available as well. In writing, where features include font size, capitalization, lines, spatial organization, and the like, the author has different and somewhat restricted means. Text “typically has to stand on its own entirely, without any help from the situation, and therefore has to supply all the necessary information explicitly” (Stubbs 1980, p. 108). In the case of a personal ad, the appealing effect of the advertiser should be much higher than in face to face communication as it would occur, for example, on a date.

Finally, Crystal (2005) addresses the origins of the two language modalities. Speaking primarily came into being to enable face-to-face communication in small groups, for example to organize hunting. Writing facilitates communication on a larger scale, overcoming locational and chronological boundaries (see also de Saussure, 1916/1983; Pinker, 2002; Bickerton, 1981). Indeed, the most basic difference between speech and writing would be their usage due to purpose, context, and subject (Chafe & Danielewicz in Horowitz & Samuels 1987). The context in which a text is composed is completely irrelevant to its production; it only becomes relevant at the moment of reading (Nystrand in Horowitz & Samuels 1987, p. 206). The author has to plan on the impression to be made by the texts, on its communicative purpose. “[T]he choice of a particular medium is normally determined by the social function of the communication” (Stubbs 1980, p. 108; cf. Horowitz in Horowitz & Samuels, 1987, p. 121). Accordingly, this which medium to use for which function is more often than not prescribed via societal norms. Chafe and Danielewicz (1987) posit that “language [is] related to the social interaction which is natural to speaking, as contrasted with the social isolation which is inherent to writing” (as cited in Horowitz & Samuels 1987, p. 87). By definition, however, any form of communication is social (cf. Merriam-Webster 2010, “social”, 3). Consequently, the written medium is social as well.

Also, with the selection of one modality come certain implications and conventions. For instance, legal contracts are always in writing while small talk is, except maybe for Internet chatting, bound to the spoken medium.

Sending an invoice is more binding than asking for money over a casual cup of coffee. The same is true for publicizing a personal, which is less daring than taking part in, say, a speed-dating event. While the data used in this study is taken from the Internet, it will be treated as printed material. The web genre criteria do not fit since the advertisements are only digitalized versions of their print publication and have not been designed specifically for the Internet. The publishing of a personal implies that either (a) the person is too shy to ask somebody out, or (b) the person has not been successful in face-to-face communication in general. They are clearly instances of written communication and not partly speech put into writing as it would be in computer-mediated communication (CMC). Moreover, identity performance is at its best in the personals when considering Chafe and Danielewicz's (1987) statement (in Horowitz & Samuels 1987, p. 88):

[T]his whole editing process is hidden from the eventual consumer of our language to whom we can pretend that the aptness of our choices flow naturally from our pen, typewriter, or word processor.

What would come out as unstructured muttering in speech can be tuned to highest efficiency in writing. The best appearance and character traits of the author can be put into the most favorable light. The addressee will not be blinded by stuttering or unwashed hair but rather by eloquent literary quotations and creative abbreviations. A thought-through, favorable presentation of self can be expected from a personal ad. Anything falling into that category will be recognized by the reader as an unmarked choice made by the author. Personal ads are hence placed within a normative framework true to the definition of Scotton (1980, v.s.). Her characteristics of identity negotiation and performance can be transferred to this written form of communication. However, the addressee is personally unknown to the author. The role relationship of the interlocutors is of a weakly defined nature (cf. Scotton, 1980). Accordingly, exploratory choices will be made by the author. In such a case, marked statements may function as attractors and will, if the reader chooses to, not be interpreted negatively. Ultimately,



all choices made shall support the goal of finding a partner, or at least a date. The gains maxim will be prioritized before general societal norms. What can be assumed, though, is that identity performance will be at its most explicit because the self- and other-description by the author of a personal will be at least slightly exaggerated. Moreover, it will be closely connected to stereotypes since a simplified picture of both interlocutors has to be drawn in a very condensed, uni-directional communication.

### 2.7.3 Stereotypes in the Personals

[W]ith members of one's own cultural group, descriptions are constructed in conventional ways according to unspoken expectations and implicit common knowledge. The hearer is expected to infer missing information cued only by the information that is included and by the genre in which the information is presented. (Holland & Skinner, 1987, p. 78)

The newspapers supplying data for the corpus analyzed in this paper are all either of a local or regional readership. Even the *NYT* contains a locally aimed classifieds section. Hence, a normative framework shared by authors and recipients of the personals can be assumed. The notions of male and female stereotypes have been outlined by Hort, Fagot, and Leinbach. (1990), providing us with a set of gender markers expected to be present in this framework. A Personal Attributes Questionnaire (PAQ) on trait-adjectives in the contexts of personality and appearance was designed. It was completed by 400 undergraduate students “with a mean age of 20.5 years” (Hort et al., 1990, p. 201). Accordingly, the analyzed population is now around the age of forty and is thus close to the average age of the advertisers in the corpus of the year 2007. A search for the numbers adjacent to occurrences of *yr*, *year*, *years*, *age*, *old* with the distance 1 in the present corpus was conducted via XQuery (see 3.2). A further search revealed that there are no instances of *year old* or *years old*. The percentages of the respective age groups are: 18-19: 3.8%, 20-29: 6.8%, 30-39: 7.8%, 40-49: 27.2%, 50-59: 36.9%, 60-69: 11.7%, 70-81:

5.8%. Accordingly, 64.1% of the advertisers analyzed are 40-59 years old. The search was conducted over the whole content part of the personals because some might only mention their own age or the desired age in a partner, or vice versa.

Apart from reiterating that one's perception of others is more stereotypical than one would expect (cf. Hort et al., 1990, p. 200), the researchers provide two lists of stereotypical adjectives used to describe males and females (cf. also Smith, 1980):

Study 1 – Personality trait-adjectives <sup>a</sup>	
Male stereotypic	Female stereotypic
Aggressive	Warm in relations with others
Independent	Submissive
Rough	Excitable in a major crisis
Competitive	Passive
Active	Devoted completely to others
Assertive	Home oriented
Worldly	Kind
Feelings not easily hurt	Aware of needs of others
Dominant	Feels inferior
Makes decisions easily	Gentle
Self-confident	Helpful to others
Feels superior	Needs security
Stands up well under stress	Cries easily
Indifferent to the approval of others	Understanding of others
Never gives up easily	Emotional

*Table 1: Personality trait adjective list (Hort et al. 1990)*

The distribution of positive and negative personality traits is balanced on both sides. While men seem to be more dominant and active women are stereotyped as submissive and social.

Study 2 – Appearance trait-adjectives	
Male stereotypic	Female stereotypic
Tall	Fair skin
Strong	Smells good
Broad-shouldered	Graceful
Sturdy build	Delicate build
Hairy	Wears jewelry
Deep voice	Soft voice
Muscular	Wears makeup
Large-boned	Small-boned
Big hands	Small feet
Short hair	Long hair

<sup>a</sup>Adapted from the PAQ (Spence et al., 1974).

*Table 2: Appearance trait adjective list (Hort et al. 1990)*

With regard to appearance, men are stereotypically strong and evoke a protective image. Women are stereotyped as slender and fair. All of the adjectives listed in Table 1 and Table 2 give rather distinctive ideas of trademarks ascribed to men and women. These stereotypes can be coordinated well with the semantic fields introduced in 2.4.2. When sorting the items listed into the lexical fields, the major concerns of women tend to be stereotypically grouped in the areas of *emotionality/eroticism/relationship* and *social behavior/charisma*, while the choice of words for men are more dominant within the semantic field of *rationality/mental disposition*. From the assumption that identity performance is intense in the personals the fulfilling of stereotypes can be deduced. Men, for instance, would want to appear as manly as possible. Moreover, Parekh and Beresin (2001) report on a 1989 study which revealed that

physical attractiveness surpassed other qualities desired by men and was listed twice as often in men's ads as women's ads. [...] Other studies have concluded that only this male attribute [of financial security] is equally exchangeable for that of beauty in women. (p. 226; cf. also Hirschman 1987)

This suggests that women and men look for the positive aspects of the stereotypes in the studies presented by Hort et al. (1990, Tables 1+2). However, it is to be expected that both sexes are more or less equally concerned with *appearance, adequacy/compatibility, education, vocation/financial assets, and talent*. One reason for this is that advertising for a partner in the classifieds allows the expression of compatibility desires. For instance, someone who cares about their own appearance will also do so in their partner. A woman who likes to play tennis would probably also like her life partner to like doing this. As Little, Burt, and Perrett (2006) report on appearance similarity between life partners,

[co]rrelations show [...] that perceived age, attractiveness and some personality traits were similar between partners and that matching for

perceived personality occurred even when controlling for age and attractiveness of the faces. (p. 1)

Hence, looking for similarities can be expected when people can choose who they desire to be with. In contrast to the popular saying, opposites do not seem to attract. Rather, 'like seeks like' is expressed in assortative mating (Puhler, 1968; cf. also Fiore & Donath, 2005; Burley, 1983; Figueredo, Sefcek, & Jones, 2006). Accordingly, whether appearance or character traits appear in self- or other-description in a personal advertisement is irrelevant. Both occurrences prove that these aspects are relevant to the author. Moreover, an idealized yet simplified advertisement of oneself is the most encompassing and appealing way to ensure a reply. For the analysis of this form of self- and other-description, semantic fields help in identifying preferences. With these as an aid, the reality of advertiser stereotypes in the personals can be exploited. In order to enable a search for the lexical fields, however, the corpus collected for this study had to be formed and marked up first.

### 3 Empirical Part – Data Collection and Processing

There is expected to be more of a tendency to place an advertisement for a relationship in a newspaper, which is a highly anonymous environment. This is a given in overpopulated areas, and especially in large metropolitan cities. There, people come to work during the week but then leave to live in the exurbs, or commuter towns. Frequent changes in place of residence, for instance because of work or for education, also increase social detachment. Accordingly, the largest number of personal ads placed is to be expected in the megalopolises located in the United States of America with a high fluctuation in population. There, the mobility rate is high and many people leave the urban areas and commute instead. For these reasons, personal ads from US American cities with a high population density (e.g. NYC, NY) as well as cities with a wide metropolitan area (e.g. Hartford, CT) have been chosen for a specialized corpus. In order to get an encompassing impression, the sources were selected randomly but for the condition of geographical diversity. Hence, the data to be analyzed contains personal ads from *The New York Times* (NY), *The Tucson Weekly* (AZ), *The Boston Phoenix* (MA), *The Hartford Advocate* (CT), *The Washington Post* (DC), and *The Detroit Metro Times* (MI). Since American newspapers are hard to get hold of in Germany, the collected personal ads have been taken from the Internet versions of the publications. However, only those ads that were actually published in print have been included in order to guarantee a unified text type. The data collection took place from April through May, 2007. This time slot was chosen because, on the one hand, its temporal brevity allows a synchronic analysis of the data. On the other hand, when considering folk wisdom the season of spring should cause more people to seek a partner.

While the aggregated database contains a total of 507 items, only ads aimed at the opposite sex have been chosen for this study. The reason for this was to avoid non-biological gender issues (see 2.1.2). The content words (see 2.4.2) of the 433 remaining personal ads from six states with

the directions "Women seeking Men" (200) and "Men seeking Women" (233) will be examined to facilitate less regional but more US-encompassing conclusions. The slight divergence in the number of advertisements is due to the availability of data at the point of the present study. The items have been taken from a previous study (Kirchhof, 2007b). Unfortunately, the accumulation of additional personals to even out the difference in number was not possible because classifieds in general are not kept in the archives of the respective e-papers. Furthermore, a deletion of random items from the female heterosexual group was decided against in order to preserve as much material as possible for an analysis. The divergence in amount has been eliminated by eliciting percentages from the data. This way, significant conclusions can be taken from the results.

In order to learn which words have been used in the advertisements collected as well as to enable more efficient processing, the data was formally structured elements using the eXtensible Markup Language (XML). Each personal ad received a unique identifier (ID) to simplify reference for later retracing. Furthermore, the surface structure of the ads was implemented, that is the title and body. For means of clear organization and usability, the date and venue of publishing of the personal ads were also included in the labeling. In order to facilitate a search of the data by sexuality, the directions of the advertisements were also entered. An exemplification of the structural categories, here applied to Example 1 (v.s.), can be seen in Table 3:

Label	Content
Source	Hartford Advocate
Date	28.05.07
Direction	Men > Women
Title	A GOOD MAN 4 U SM, 60,
Body	creative, caring, compassionate, likes home cooking, classic rock, writing. WLTM similar lady to share talks, casual dates, friendship and fun. Possible LTR.

*Table 3: Structure of Personal Advertisement.*

This structure transferred onto all the personals with the help of an XML schema (see Appendix). The resulting corpus MATRI (Kirchhof, 2007a) is of a specialized nature. It is not balanced by general definition (cf. Dickinson, 2009) because it neither includes speech nor any other modalities or genres but print personals. However, as a genre-specific corpus, MATRI is balanced within its domain. Sub-corpora, for instance of the particular genders, sexes, or age groups are easily obtainable via the textual markup. Moreover, the audience type is shared by all the ads in this synchronic corpus. The richly annotated collection containing the personal ads collected was loaded onto a local server at Bielefeld University to be independently accessible. Since a linguistic analysis was to be conducted on the corpus, however, further preparations had to be made. Since the data was now present in XML format, the corpus was tagged with POS labels as attributes for each word. Further reasons for this action as well as the procedure itself will be explained in the following section.

### **3.1 Part-of-Speech Tagging**

The automatic extraction of linguistic information from large corpora enables rapid progress in understanding language in all modalities (Marcus, Marcinkiewicz, & Santorini, 1993, p. 313). The searchability of data increases along with the appropriateness and detail of its annotation. Accordingly, treebanks (parsed corpora) are produced.

For the annotation of the MATRI corpus with POS tags, the method of the Penn Treebank, which was also used for the Switchboard corpus (UPenn, 1999) has been adopted. Hence, the development and markup of this method in connection to corpus annotation will be outlined in this section. Based on this, the actual processing of the data used for this paper will be described in the succeeding section. The databank of the Penn Treebank project (see also UPenn, 1999) consists of more than 4.5 million documents of written and spoken American English. All of these are present as linguistic trees, that is they are “parses showing rough syntactical and semantic information” (UPenn, 1999, ¶1). Its variety of text types supplies high overall lexical density. Among the entries are data from the Library of America, Dow Jones Newswire stories and an edited version of the Brown Corpus (Marcus et al. 1993, p. 327). The latter was the first corpus ever to be annotated, that is tagged, automatically for POS. As Marcus et al. (1993) explain,

[t]he rationale behind developing such large, richly annotated tagsets is to approach 'the ideal of providing distinct codings for all classes of words having distinct grammatical behaviour' (Garside, Leech, & Sampson (1987). (p. 314)

And, via these grammatical attributes, the lexical level of the items tagged can be reached. Corpora marked up in such a way are often used in natural language processing (NLP). Moreover, they are great resources for sociolinguistic analysis. The Brown Corpus annotation schema was redundant in terms of sub-classifications, for instance temporal and causal



adverbs and five different forms of main verbs (Marcus et al., 1993, p. 313; cf. Garside et al., 1987). For the Penn Treebank tagset, lexical as well as syntactic information was combined, eliminating such redundancy. This means that the type of qualifiers, for example, will be annotated with a synergy of both categories. However, the corpus allows more explicitness as well if desired (Marcus et al. 1993, p. 314). In comparison to the Brown Corpus, the reduction of possible tags also contributes to the consistency in tagging. The direct integration of syntactical information into the Penn tagset has a similar effect. *One*, for example, “is [now] tagged as NN (singular common noun) rather than as CD (cardinal number) when it is the head of a noun phrase” (Marcus et al., 1993, p. 316). This may be irrelevant for tagging such lexically dense texts as personal advertisements, where no real syntax is present. Contrary to this, in more complex academic or literary texts, this function might be especially desirable.

As can be seen in the example given below, the Penn annotation is rather idiosyncratic as it builds on the conventions of the Brown Corpus and some aspects have been modified. Determiners, for example, are tagged with 'DT', adverbs with 'RB', and adjectives with 'JJ' (for full list see Appendix). In Example 8, the title of a personal ad has been annotated with the Penn Treebank tagset. Here, the two-letter 'TAG' attributes of the 'POS' elements give information on the parts of speech:

```
<Matrimonial ID="dle39">
  <Title>
    <POS TAG="DT" LEMMA="a">A</POS>
    <POS TAG="RB" LEMMA="very">VERY</POS>
    <POS TAG="JJ" LEMMA="shy">SHY</POS>
    <POS TAG="" LEMMA="">.</POS>
    <POS TAG="" LEMMA="">.</POS>
    <POS TAG="" LEMMA="">.</POS>
  </Title>
</Matrimonial>
```

**Example 8:** Title “A very shy ...” of a personal ad tagged for POS.

During the POS annotation of the Penn Treebank, two stages were applied. First, an automatic tagging took place. PARTS, a stochastic algorithm, was used to parse the corpus (see Marcus et al., 1993, p.317). After the result had been automatically tokenized, it was mapped onto the Penn Treebank tagset. An error rate of 7-9% was common in the early development stages, but further research has continuously brought improvement. New taggers using stochastic and rule-driven methods now only return 2-6% of errors. Another important modification is that the tagger is now directly based on the Penn Treebank tagset. It is not anymore an alignment of a modified Brown Corpus set with the Penn one (Marcus et al., 1993). In the second stage of the Penn corpus creation, semi-manual corrections were done. Fortunately, the second process is displayed in confusion matrices, providing grounds for future improvement. Furthermore, a comparison of manual-only tagging with a correction process of automatized tagging showed that the method described above was more than twice as efficient and only half as error prone (Marcus et al. 1993, p. 319). Finally, the complete Penn Treebank was tagged with an error rate as low as 3%, making the tagging method highly usable for corpus POS annotation of any kind.

Another part of the Penn tagset is the automatized syntactic bracketing. Fidditch (Hindle 1983; 1989), another parser, is implemented as well (as cited in Marcus et al. 1993, p. 320). Not only does it have a low error rate but it also returns separate sections when in doubt, so-called 'null elements'. This means that the parser provides the user only with fully functional structures, albeit allowing for manual integration of the separated ones (Marcus et al. 1993). And, again, a manual stage is appended to the automatized process. The efficiency is undeniable since “[a] parsed subcorpus of over 1 million words was recently proofread at an average speed of approximately 4,000 words per annotator per hour” (Marcus et al. 1993, p. 323). The usage of this additional tagging procedure was, however, omitted for the present study because of a research focus on the content level of the personals.

Nevertheless, there are some limitations to the Penn annotation schema. For example, due to the context-free nature of the corpus, argument and adjunct relationships are prone to errors, which is also true for certain predicate-argument structures (Marcus et al. 1993, p. 329). This, however, will not be detrimental to the current study since rarely any regular syntax is present at all in personal ads. Another flaw of the Penn tagset may be that it is not detailed enough, but improvements are in progress. For regular analyses, however, and definitely for the tagging of printed personals, it is fully sufficient.

### **3.2 Terminology Extraction and Data Processing**

In order to learn about the topics that are important to the issuers of the personal ads, only content words, that is nouns, adjectives, verbs, adverbs, as well as their respective variations have been taken into account. For that purpose, the corpus was annotated further according to POS with the Penn Treebank POS tagger. After a transformation into XML format, the data was loaded into a Tamino database. This type of databank, in contrast to MS Excel or SQL, is not tabular but has a tree structure. This facilitates a more linguistic and more structured approach to the data. The personal ads in XML format can now be examined by using the request language XQuery, which is utilizing, among other methods, XPath and regular expressions. With the help of structural as well as semantic restrictions, XQuery navigates within the XML tree structure (cf. W3C, 2007 on <http://www.w3.org/TR/xquery/>). The word lists and statistics to be presented in the next section have been extracted from the database with variations of Query 1 (see Appendix, including comments), revealing their deep structure. The method of word elicitation will be presented in the following. Before proceeding to the deep structure of the data and to the presentation of the results, however, a short explanation about the choice of categories in labeling the personal ads in this paper will be given.

As has been mentioned above, in order to elicit relevant information about the relationship preferences of the advertisers, the content words of the textual items have to be analyzed. "Apart from possible problems with classification in terms of parts of speech, this does not preclude idioms as special lexical elements from being members of lexical fields" (Lutzeier, 2006, p. 80). However, the open source tool *TreeTagger* (Schmid, 1997) makes it possible to automatically tag all parts of speech in an English-language text according to the Penn annotation schema (see also <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>). An interface with copy-and-paste function makes tagging a matter of seconds. After being parsed with the *TreeTagger*, the ads are annotated richly enough to

be of use for deeper linguistic research. While all items in the MATRI corpus have been tagged, only the content-words were processed further.

After the transformation of the enhanced data into XML, the queries on the corpus could be restricted by POS, and thus by distinct markers of the linguistic profiles of all authors. As has been argued in 2.4.2, this encompasses the following word classes: adjectives (JJ, JJR, JJS), nouns (NN, NNS), proper nouns (NNP, NNPS), pronouns (PRP, PP\$), adverbs (RB, RBR, RBS), and verbs (VB, VBD, VBG, VBN, VBP, VBZ) (see Appendix for detailed explanation). With an XQuery containing the XPath to the POS attributes of all words, all lexical items of one or more word classes can be elicited from the corpus. With further search conditions using the XML structure, for instance “Women seeking Men” as direction or the source “New York Times”, a multitude of questions can be asked and answered. Thus, the frequency of adjectives, for instance, can be elicited, specified as to which ones are used the most and least, by whom and where.

During the word elicitation process, verbs were excluded from further analysis. A request for all verbs in the corpus resulted in the output of table 2, only returning forms of *be* and auxiliaries and no semantically laden verbs.

to be				other auxiliaries			
is	50	Is	5	can	19	Must	2
be	38	Be	3	will	15	may	2
am	24	been	3	must	10	could	1
are	22	Are	2	would	7	might	1
being	7			Would	5	should	1

*Table 2: All verbs in the present corpus.*

Since both, copula and auxiliaries belong within the group of function words, they do not have content value. This lead to the decision to disregard them in the continuing analysis of the language of personals in this paper.

### **3.3 Elicitation of Lexical Words**

By modifying Query 1, the result can be shown in alphabetical order, sorted by length and number of appearance, and restricted to location, direction, and other categories (see Appendix). Furthermore, the distribution of POS by frequency can be retrieved. On top of that, since the majority of advertisements mention the age of the author in numerals, this influence on their language can also be taken into account for analysis (see 2.7.3).

The content words were elicited from the data to make a socio-cultural analysis of the advertisements possible. This procedure was only conducted on the body of the personal ads. The inclusion of the headings, especially with their tendency for abbreviations and quotations, would have denaturalized the language used by the authors. Again, a query was used to return a list of all content words found in the corpus. The result was then sorted by word frequency, allowing ratings within and across word classes. First, the overall POS numbers were elicited in order to form a basis for further comparison (see Table 3). Second, the respective ratings of POS across partner-search directions were extracted. This enabled a contrastive examination of general tendencies compared with the data taken from heterosexual men and women, respectively. Following the elicitation, the assignment of words to the semantic fields had been done manually, and, as a consequence, can be considered highly subjective (cf. 2.4.2). Nonetheless, the references to Gottburgsen (1995, pp.272-273) provided some consistency. The frequencies of the different semantic fields used by the advertisers of the personals will hint at the relevance of these topics to them. It will be analyzed whether the preferences fit the expectations formed in 2.7.3. Most importantly, though, the subsistent gender preferences with regard to linguistic expression will be revealed.

### 3.4 Results

The whole MATRI corpus includes 433 ads with heterosexual direction, which contain 1968 distinct lexical words. This set consists of 777 different adjectives (39.5%), 160 adverbs (8%), and 1031 nouns (52.5%). The ten most frequently used content words in all cross-sex-directional personals in the corpus are as follows: *Attractive* (91), *good* (85), *fun* (70), *life* (66), *movies* (66), *friendship* (62), *share* (56), *intelligent* (55), *music* (54), and *fit* (53) (see Table 3). Due to their obligatory occurrence of the tokens *man* (70), *woman* (70), and *BOX* (53), these have been excluded from the ratings.

Within the 200 personals of “Men seeking Women”, 1126 distinct content words were found, which makes up for 55% of the whole corpus. These are composed of 424 adjectives (37.7%), 100 adverbs (8.9%), and 602 nouns (53.5%). The percentages in parentheses are calculated in relation to all content words in this directional category. The ten most frequently used lexical words directed from men towards women in the corpus are as follows: *Attractive* (44), *good* (42), *friendship* (33), *fun* (33), *life* (33), *share* (28), *movies* (25), *music* (25), *romance* (25), and *intelligent* (23). These are compared to the overall ratings in Table 3. Again, instances of *woman* (52) have been excluded because of their low content value. *More* (40), which in this context is high in frequency as a modifier, does not supply lexical content either and was hence also omitted.

In contrast to this, the 233 personal ads with the direction “Women seeking Men” show 1405 different content words, that is 68.3% of those in the whole heterosexual MATRI data. These are composed of 589 adjectives (42%), 109 adverbs (7.8%), and 707 nouns (50.3%). The ten most frequently used lexical items in this group are: *Attractive* (47), *good* (43), *movies* (41), *fun* (37), *interests* (33), *life* (33), *intelligent* (32), *fit* (31), *friendship* (29), and *music* (29). Tokens of *man* (56) and *BOX* (36) were disregarded (v.s.).

distinct words	five most frequent
<b>all (2057)</b>	<i>attractive</i> (113), <i>good</i> (85), <i>fun</i> (70), <i>life</i> (66), <i>movies</i> (66)
<b>nouns (1031)</b>	<i>fun</i> (70), <i>life</i> (66), <i>movies</i> (66), <i>friendship</i> (62), <i>share</i> (56)
<b>adjectives (777)</b>	<i>attractive</i> (113), <i>good</i> (85), <i>intelligent</i> (55), <i>slim</i> (49), <i>honest</i> (45)
<b>adverbs (160)</b>	<i>more</i> (70), <i>very</i> (40), <i>long</i> (29), <i>financially</i> (28), <i>pretty</i> (16)

Table 4: Content word frequency in heterosexually directed personal ads.

Having grouped the 83 (STDV: 3.0) most frequent content words of each category into the semantic fields adapted from Gottburgsen (1995), the overall percentages as seen in Illustration 3 can be derived. The percentages refer to word numbers within the groups. Here, the most preferred semantic fields by heterosexual men and women, respectively, are set against the overall percentages. Since the difference in percentage was too small across categories, a test of significance was deemed unnecessary.

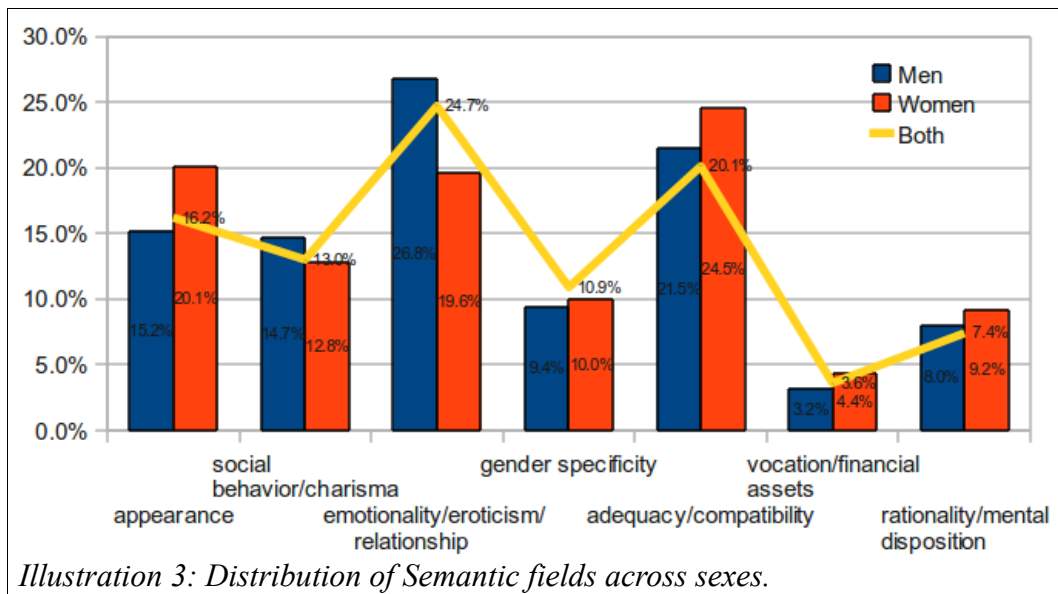


Illustration 3: Distribution of Semantic fields across sexes.

The favorites across sexes are within the fields of *emotionality/eroticism/relationship* and *adequacy/compatibility*. In this, women have the highest ratings in the latter category (24.5%), followed by *appearance*



(20.1%) and *emotionality/eroticism/relationship* (19.6%). Men, in contrast, are slightly higher than the cross-sex ratings of and lead in the words belonging to *emotionality/eroticism/relationship* (26.8%). No lexemes belonging within the fields of *education* or *talent* were used in these most frequent content words by either sex, nor were there any in the MATRI corpus.

The largest age group in the data is made up of those being between 40 and 59 years old (64.1%). Accordingly, the overall word frequencies are, on the one hand, strongly influenced by this population. On the other hand, this majority is highly representative of the authors of personal ads in the MATRI corpus in general. Hence, the words statistics of men and women in their forties and fifties have been elicited to demonstrate the dominant tendencies. In the category “Women seeking Men”, the ten most frequently used lexical items are: *Attractive* (47), *good* (43), *movies* (41), *fun* (37), *interests* (33), *life* (33), *more* (30), *travel* (29), *relationship* (27), and *humor* (26). Instances of *BOX* or *man* do not figure in here. This makes for a high concentration in the semantic fields of *emotionality/eroticism/relationship* (38.4%) and *adequacy/compatibility* (48%) overall in heterosexual women aged 40-59 years in the present data. *Appearance* factors in with 13.6%.

The men of the same age group have the following most frequently used lexical items, exclusive of *woman*: *Attractive* (44), *good* (42), *more* (40), *friendship* (33), *fun* (33), *lady* (29), *movies* (25), *music* (25), *romance* (25), and *intelligent* (23). Accordingly, their dominant semantic fields are also *emotionality/eroticism/relationship* (44%) and *adequacy/compatibility* (42.3%). However, the overall preferences differ. *Appearance* is always a factor (13.8%) as well, but *attractive* is the only lexical instance in the top forty content words.

## 4 Analysis and Discussion

Resulting from a combination of the semantic fields defined in 2.4.2. and the stereotype predictions of the PAQ reported by Hort et al. (1990), certain predictions of the preferences of men and women advertising in the personals of US American print newspapers were made. Suggested by the literature on gender stereotypes, women were expected to favor the semantic fields of *emotionality/eroticism/relationship* and *social behavior/charisma*, while men were assumed to be more dominant in *rationality/mental disposition*. In addition to that, a rather balanced usage of content words belonging to the fields of *appearance*, *adequacy/compatibility*, *education*, *vocation/financial assets*, and *talent* was predicted.

Since neither *education* nor *talent* could be attested for in the present corpus, the formation of these fields might have been faulty. It is possible, however, that the field *vocation/financial assets* compensates for this lack. This is supported by the fact that the latter category, along with the areas of *appearance* and *adequacy/compatibility*, is indeed among the most equally-often mentioned topics across sexes, although not with a high percentage (~4%). The purpose of finding the perfect match explains the high ratings of the *adequacy/compatibility* field, exceeding 20% within both sexes, respectively. The anticipated gendered preferences, however, have been refuted by the data. Men, for instance, not only used a low percentage of words concerning *rationality/mental disposition*, they were even surpassed by the women in this field. Furthermore, they were shown to use 3-5% more from the semantic fields of *emotionality/eroticism/relationship* and *social behavior/charisma* than females, nullifying the predictions made for the favorite areas of women. The ratings suggest that women attach most importance to compatibility (24.5%), appearance (20.1%), and relationship design (19.6%) in finding a match. These areas are also the three most valued issues with men, with 26.8% for relationship design, 21.5% for compatibility, and 15.2% for appearance. The shared favorites indicate that

both males and females aim at the same goal, namely finding a partner, and thus follow the 'rules' for reaching a successful result. The opposing frequencies of previously gender-stereotyped semantic fields suggest that the advertisers might aim at the other sex's stereotypes in order to influence their success rate in finding a partner. However, this cannot be researched further with the present corpus as the subjects cannot be questioned about their intentions.

What has been determined by the present study, however, are the stereotypical authors of US American personal ads: Heterosexual men and women in their 40s and 50s. As has been discussed in 2.2, people move homes rather frequently because of changes in work place. People in the middle of their lives, especially from the USA, are likely to have moved as well as changed jobs more than once already. Another reason for the high number of mid-lifers placing personal ads are the high divorce rates in the US (cf. 2.2). According to the Forest Institute of Professional Psychology in Springfield (MS), "50% percent of first marriages, 67% of second and 74% of third marriages end in divorce" (<http://www.divorcerate.org/>). The emphasis in the desire for *emotionality/eroticism/relationship* (44%), that is for a more harmonic and lasting partnership, are nearly twice as high as for the heterosexual population in the corpus in total (24.7%). Furthermore, the importance of *adequacy/compatibility* in a relationship at the age of 40-59 more than doubles from 20.1% overall to 42.3%. While already only figuring in with 16.2% on average, attractiveness in a new partner is cared even less about by men (13.8%) as well as by women (13.6%) in this age group. The average US American, middle-aged single is looking for harmony and stability in a relationship while other aspects become peripheral. What is more, going to the 'movies' is their most preferred recreational activity, demonstrating a preference for going out. This group is also the one most frequently relying on the print medium to find a partner (64.1%).

Finally, to pay tribute to general stereotyping, with regard to linguistic colorfulness, women in general exceed men by 13.3%. At least in terms of identity performance in printed personal ads, women are of a more

'talkative' nature than men. However, they do not put much importance on finances, the topic of which is negligible in general in the present corpus. The hypothesis that the lexical inventory used is gender-neutral has been falsified by the MATRI corpus across all groups. This also contributes to the common stereotype that men are more factual than the females. Apart from these two aspects, however, the expectation that sociocultural stereotypes are not significantly dominant in the advertisements has been verified by the data.

A less subjective grouping of the fields would enhance the representativeness as well as the objectivity of the study. Unfortunately, semantics cannot be freed from subjectivity in the context of analyzing written, unidirectional communication. Another expansion of the study of gender preferences in partner finding to same-sex ads and with regard to the number of analyzed items could serve to further clarify the distribution of semantic fields.

## 5 References

### *Literature*

- Allwood, J. (2006). Cognitive science 'consciousness, thought and language'. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2<sup>nd</sup> ed.) (pp. 44-53). Oxford: Elsevier.
- Bartsch, R.A., Burnett, T., Diller, T.R., & Rankin-Williams, E. (2000). Gender representation in television commercials: Updating and update. *Sex Roles*, 43, 735-743.
- Beauman, F. (2010). *Still looking for Mr. Right*, Feb 10, 2010. Retrieved from <http://www.thedailybeast.com>
- Bem, S.L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42, 155-162.
- Benwell, B. and E. Stokoe (2006). *Discourse and identity*. Edinburgh: Edinburgh UP.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: California UP.
- Best, D. L. (2004). Gender stereotypes. In Ember, C.R., & Ember, M. (Eds.), *Encyclopedia of sex and gender: Men and women in the world's cultures* (pp. 11-23). New York, NY: Kluwer.
- Bickerton, Derek (1981). *Roots of Language*. Ann Arbor, MI: Karoma Publishers.
- Bloomfield, L. (1933). *Language*. New York, NY: Henry Holt.
- Bruthiaux, P. (1996). Functional variation in the language of classified ads. *Perspectives: Working Papers of the Department of English, City Polytechnic of Hong Kong*, 6(2), 21-40.
- Burley, N. (1983). The meaning of assortative mating. *Ethology and Sociobiology*, 4, 191-203.
- Castelo-Branco, C., Huezco, M.L., & Lagarda, J.L. (2008). Definition and diagnosis of sexuality in the XXI century. *Maturitas*, 60, 50-58.
- Cheshire, J. (1998). The relationship between language and sex in English. In Linn, M. D. (Ed.), *Handbook of dialects and language variation* (pp. 407-418). San Diego, CA: Academic Press.
- Chomsky, Noam (1957). *Syntactic structures*. The Hague/Paris: Mouton
- Coates, J. (1986). *Women, men and language*. London: Longman.
- Crystal, D. (2005). *How language works*. Woodstock, NY: Overlook Press.
- Crystal, D. (2003). *A dictionary of linguistics and phonetics* (5th ed.). Oxford: Blackwell.
- Day, K., Gough, B., & McFadden, M. (2004). 'Warning! Alcohol can seriously damage your feminine health': A discourse analysis of recent British newspaper coverage of women and drinking. *Feminist Media Studies*, 4(2), 165-185.

- Dickinson, M. (2009). Basics of corpus linguistics [PDF document]. Retrieved from <http://jones.ling.indiana.edu/~mdickinson/09/615/slides/02-basics.pdf>
- Divakaruni, C. B. (2000). Uncertain objects of desire. *The Atlantic Monthly*, 285(3), 22-27.
- Dixon, J.A., & Foster, D.H. (1997). Gender and hedging: from sex differences to situated practice. *Journal of Psycholinguistic Research*, 26, 89-107.
- Edwards, R., & Hamilton, M. A. (2004). You need to understand my gender role: An empirical test of Tannen's model of gender and communication. *Sex Roles*, 50, 491-504.
- Ember, C. R., & M. Ember (Eds.). (2004). *Encyclopedia of sex and gender: Men and women in the world's cultures*. New York; London: Kluwer.
- Figueredo, A. J., Sefcek, J. A., & Jones, D. N. (2006). The ideal romantic partner personality. *Personality and Individual Differences*, 41(3), 431-441.
- Fiore, A.T., & J.S. Donath (2005). Homophily in online dating: When do you like someone like yourself?. Short paper presented at: ACM '05: *Computer-Human Interaction*. Portland, OR.
- Frable, D. E. S. (1989). Sex typing and gender ideology: Two facets of the individual's gender psychology that go together. *Journal of Personality and Social Psychology*, 56, 95-108.
- Gahl, S., Garnsey, S. Matzen, L. & Fisher, C. (2006). 'That sounds unlikely': Phonetic cues to garden paths. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.
- Garside, R., G. Leech, & G. Sampson (Eds.). (1987). *The computational analysis of English: A corpus-based approach*. London: Longman.
- Geeraerts, Dirk (2002). The scope of diachronic onomasiology. In Agel, V., Gardt, A., Hass-Zumkehr, U., & Roelcke, T. (Eds.), *Das Wort. Seine strukturelle und kulturelle Dimension. Festschrift für Oskar Reichmann zum 65. Geburtstag* (pp. 29-44). Tübingen: Niemeyer.
- Gibson, J. (2010). Sexless in the city: a gender revolution. *The Sydney Morning Herald*, 12 Mar, 2010. Retrieved from <http://www.smh.com.au/>
- Gilbert, L. A. (1985). Dimensions of same-gender student faculty role-model relationships. *Sex Roles*, 12, 111-123.
- Gottburgsen, A. (1995). Zur sprachlichen Inszenierung von Geschlecht – doing gender in Kontaktanzeigen. *Zeitschrift für germanistische Linguistik*, 23, 257-283.
- Gramley, S. (2001). *The Vocabulary of World English*. London: Hodder Arnold.
- Gramley, S., & V. Gramley (Eds.). (2008). *Bielefeld introduction to applied linguistics*. Bielefeld: Aisthesis.
- Gramley, S., Pätzold, K.-M. (2004). *A survey of modern English* (2nd ed.). London: Routledge.
- Grice, H.P. (1975). Logic and conversation. In Cole, P., & Morgan, J.L. (Eds.), *Speech acts* (pp. 41-58). New York, NY: Academic Press.

- Groom, C.J., & Pennebaker, J.W. (2005). The language of love: Sex, sexual orientation, and language use in online personal advertisements. *Sex Roles*, 52, 447-461.
- Hale, M. (1995). *Personal Ads: A career woman's approach*. Del Mar, CA: Talking Heart's Press.
- Halliday, M.A.K. (1989). *Spoken and written language*. Oxford: Oxford UP
- Hentges, B., Bartsch, R., & Meier, J. (2007). Gender representation in commercials as a function of target audience age. *Communication Research Reports*, 24, 55-62.
- Hirschman, E. (1987). People as products: analysis of a complex marketing exchange. *Journal of Marketing*, 51, 98-108.
- Hoffman, R. M., & Borders, L. D. (2001). Twenty-five years after the Bem Sex-Role Inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development*, 34, 39-55.
- Holland, D., & D. Skinner (1987). Prestige and intimacy. The cultural models behind Americans' talk about gender types. In Holland, D., & Quinn, N. (Eds.), *Cultural models in language and thought* (pp. 78-111). New York: Cambridge UP.
- Hopper, R. (1972). Overgeneralization as a learning strategy in communicative development. *Western Speech*, 36(1), 47-52.
- Horowitz, R., & S.J. Samuels (Eds.). (1987). *Comprehending oral and written language*. San Diego, CA/London: Academic Press.
- Hort, B.E., B.I. Fagot, & Leinbach, M.D. (1990). Are people's notions of maleness more stereotypically framed than their notions of femaleness?. *Sex Roles*, 23, 197-212.
- Kaneko, M. (2010). Some men swapping pants for dresses. *The Japan Times Online*, March 24, 2010. Retrieved from <http://search.japantimes.co.jp>
- Kilman, L. (2005). Newspapers: 400 years young!. *World Association of Newspapers*, 26 Feb, 2010. Retrieved from <http://www.wanpress.org/article6476.html>
- Kirchhof, C. (2007b). Supplementary Studies on Matrimonials. In Gramley, V. (lect.), *Getting Married, Having Children and Dying in English*, WS2006/07. Bielefeld: Bielefeld University.
- Krueger, J. I., Hall, J. H., Villano, P., & Jones, M. C. (2008). Attribution and categorization effects in the representation of gender stereotypes. *Group Processes Intergroup Relations*, 11(3), 401-414.
- Laqueur, Thomas (1990). *Making sex: Body and gender from the Greeks to Freud*. Cambridge, MA: Harvard UP.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Little, A. C., Burt, D. M., & Perrett, D. I. (2006). Assortative mating for perceived facial personality traits. *Personality and Individual Differences*, 40, 973-984.

- Loxton, Edward. (2007). Thailand's 'third sex' seeks legal recognition. In Horne, N. (Ed.), *The First Post*, May 17, 2010. London.
- Lutzeier, P. R. (2006). Lexical fields. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2 ed.) (pp.79-87). Oxford: Elsevier.
- Manring, M. M. (1998). *Slave in a box: The strange career of Aunt Jemima*. Charlottesville, VA: Virginia UP.
- Marcus, M.P., M.A. Marcinkiewicz, & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313-330.
- Marks, Susan (2007). *Finding Betty Crocker: The secret life of America's First Lady of food*. Minneapolis, MN: Minnesota UP.
- Marshall, A. (1994). *Classified love: a guide to understanding personal ads*. n.p.: Marshall Pub.
- McGarty, C., V. Y. Yzerbyt, & R. Spears (Eds.) (2002). *Stereotypes as explanations: The formation of meaningful beliefs about social groups*. Cambridge: Cambridge UP.
- Miller, G. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38, 11.
- Nair, R. B. (1992). Gender, genre, and generative grammar: Deconstructing the matrimonial column. In Toolan, M. (Ed.), *Language, text and context* (pp. 227-254). London: Routledge.
- Parekh, R., & E.V. Beresin (2001). Looking for love? Take a cross-cultural walk through the personals. *Academic Psychiatry*, 25(4), 223-233.
- Pinker, Steven (2000). *The Language Instinct: How the Mind Creates Language*. New York: Harper Perennial Modern Classics.
- Saussure, de, Ferdinand (1983). *Course in general linguistics*. (Roy Harris, Trans.). Chicago, IL: Open Court. (Original work published 1916).
- Schmid, Helmut (1997). Probabilistic part-of-speech tagging using decision trees. In D.H. Jones and H.L. Somers Eds.), *New methods in language processing* (pp. 154–164), London: UCL Press.
- Scotton, C. M. (1980). Explaining linguistic choices as identity negotiations. In Giles, H., Robinson, W.P., & Smith, P.M. (Eds.), *Language: Social psychological perspectives* (pp. 359-366). Oxford: Pergamon Press.
- Segal, E.S. (2004). Cultural constructions of gender. In Ember, C.R., & Ember, M. (Eds.), *Encyclopedia of Sex and Gender: Men and women in the world's cultures* (pp. 3-10). New York, NY: Kluwer.
- Singh, I. (1999). Language and ethnicity. In L. Thomas, & S. Wareing (Eds.), *Language, society and power: An introduction* (pp. 83-97). London: Routledge.
- Smith, J., Waldorf, A., & D. Trembath (1990). Single white male looking for thin, very attractive.... *Sex Roles*, 23(11/12), 675-685.
- Smith, P. M. (1980). Judging masculine and feminine social identities from content-controlled speech. In Giles, H., Robinson, W. P., & Smith, P. M. (Eds.), *Language: Social and psychological perspectives* (pp.). New York: Pergamon.



- Spuhler, J. N. (1968). Assortative mating with respect to physical characteristics. *Eugenics Quarterly*, 15, 128–140.
- Stubbs, M. (1980). *Language and Literacy: The Sociolinguistics of Reading and Writing*. London: Routledge, & Kegan Paul.
- Stubbs, M. (1983). *Discourse Analysis: The sociolinguistic analysis of natural language*. Oxford: Blackwell.
- Tannen, D. (1982). Oral and literate strategies in spoken and written language. *Language*, 58, 1-21.
- Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: Morrow.
- Tannen, D. (1994). *Gender and Discourse*. New York, NY: Oxford UP.
- Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge UP.
- Merriam-Webster, Incorporated (1994). *Webster's new encyclopedic dictionary* (2<sup>nd</sup> pr). Cologne: Könemann.
- Wierzbicka, A. (1999). 'Universals of colour' from a linguistic point of view. *Behavioral and Brain Sciences*, 22, 724-725.
- Woolrich, C. (as Irish, W.). (1947). *Waltz into Darkness*. (3<sup>rd</sup> pr.). Philadelphia, PA: Lippincott.

### **Websites**

- Bounty (UK) Ltd (2010). *Baby Gender Predictor*. Retrieved May 25, 2010, from <http://www.bounty.com/organiser-tools/gender-guesser>
- F.A.Z. Electronic Media GmbH. (2010). *FAZ.net*. Retrieved May 25, 2010, from <http://anzeigen.faz.net/>
- Harper, D. (2010). *Online Etymology Dictionary*. Retrieved May 25, 2010, from <http://www.etymonline.com/>
- Merriam-Webster, Incorporated. (2010). *Merriam-Webster Online Dictionary*. Retrieved May 25, 2010, from <http://www.merriam-webster.com/>
- Oxford Word and Language Services. (2010). *Oxford English Dictionary*. Retrieved May 25, 2010, from <http://www.askoxford.com/>
- UPenn. (1995). The Penn Treebank Project. Retrieved May, 29, 2010, from <http://www.cis.upenn.edu/~treebank/>

### **Databanks**

- Buckmaster, J. (CEO). (2008). *Craigslist Winston-Salem*. Retrieved January 6, 2008, from <http://winstonsalem.craigslist.org/>
- Diederich, Francis (President, & General Manager) (2007). *Newspapers.com*. Retrieved May-June, 2007, from [www.newspapers.com](http://www.newspapers.com)
- Kirchhof, C. (2007a). *MATRI. Tamino database*. Retrieved April-May, 2007, from *The New York Times* (2007), *Craigslist* (2007) and *Newspapers.com* (2007). Hosted on <http://olive.spectrum.uni-bielefeld.de/tamino/MATRI>

New York Times Company, The (2007). *The New York Times*. Retrieved May-June, 2007, from <http://www.nytimes.com/>

### ***Digital Media***

"Betty Crocker advertisement, 1971." JPEG. General Mills, Inc. Advertising Is Good For You, Zargaj, P. (Ed.). 25 May 2010. <[http://pzrservices.typepad.com/vintageadvertising/images/2008/09/28/1971\\_betty\\_crocker\\_ad.jpg](http://pzrservices.typepad.com/vintageadvertising/images/2008/09/28/1971_betty_crocker_ad.jpg)>

"Vplus advertisement." GIF. Brauerei C. & A. Veltins. 100Partnerprogramme.de. 25 May 2010. <<http://www.100partnerprogramme.de/uploads/pics/vplus.gif>>

Finestra, C. (Producer/Writer). (1991-1999). *Home Improvement*. New York, NY: American Broadcasting Company (ABC).

## 6 Appendix

### 6.1 Query 1

```
declare namespace ft="http://www.w3.org/2002/04/xquery-operators-text";
let $corpus := collection("MATRI")/MatriCollection/Matrimonial/Content/POS
<!-- Going down in the data structure tree to the level of POS -->
for $words in distinct-values
<!-- selecting only unique items -->
(for $i in $corpus
where $i/@TAG="JJ"
<!-- Selection of unique adjectives (Penn Annotation Schema: "JJ" = adj.) -->
<!-- restricting the search to the category of Men looking for Men -->
and $i/../../Direction="Men > Men"
return $i)
let $sorted := count
<!-- counting the number of tokens of the unique lexical items (types) to elicit
their frequency in the corpus →
(for $frequenz in $corpus
where $frequenz=$words
return $frequenz)
<!-- order them by frequency from 1 to n -->
order by $sorted descending
return
<!--The output is a table with two columns, i.e. one for lexemes and their
frequency in the corpus, respectively.-->
<tr>
  <td>{$words}</td>
  <td>{$sorted}</td>
</tr>
```

Query 1.

### 6.2 XSchema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns="http://www.w3.org/2001/XMLSchema"
  xmlns:Matrimonial="matri.xsd">
  <xs:element name="Matrimonial" minOccurs="1">
    <xs:complexType>
      <xs:attribute name="ID" type="xs:ID"/>
      <xs:sequence>
        <xs:element name="Source" type="xs:anyURI"/>
        <xs:element name="Date" type="xs:date"/>
        <xs:element name="Direction" type="xs:string"/>
        <xs:element name="Title" type="xs:string"/>
        <xs:element name="Content" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

### **6.3 Words grouped by semantic fields and personal ad category.**

#### **6.3.1 All lexical items used by heterosexuals grouped by semantic field**

**appearance (405):** attractive (113), slim (49), tall (35), active (27), hair (27), handsome (26), brown (24), eyes (22), athletic (21), blue (20), healthy (15), blond (13), petite (13).

**social behavior/charisma (324):** good (85), fun (70), kind (37), humor (35), active (27), easygoing (21), nice (18), outgoing (18), open (13)

**emotionality/eroticism/relationship (617):** life (66), friendship(62), share (56), love (48), relationship (48), loving (45), romance (43), romantic (32), times (27), companionship (25), warm (25), time (23), term (18), children (16), affectionate (15), dates (15), compassionate (14), long- (14), dinner (13), heart (13), open (13), passionate (13)

**gender specificity (272):** man (70), woman (70), lady (43), gentleman (29), male (24), female (21), guy (15)

**adequacy/compatibility (503):** movies (66), music (54), interests (53), fit (53), travel (47), similar (35), smoker (27), theater (23), Jewish (20), conversation (19), 60s (18), 40s (16), activities (16), children (16), concerts (14), art (13), same (13)

**education (0):** n/a **talent (0):** n/a

**vocation/financial assets (89):** secure (34), professional (32), successful (23)

**rationality/mental disposition (184):** intelligent (55), honest (45), down-to-earth (29), sincere (24), smart (16), healthy (15)

#### **6.3.2 All lexical items used by heterosexual women, grouped by semantic field**

**appearance (283):** attractive (61), slim (29), blond (25), tall (23), eyes (18), active (17), hair (15), blue (13), brown (13), handsome (10), well- (9), athletic (9), petite (9), slender (9), dark (8), good- (8), beautiful (7)

**social behavior/charisma (181):** good (43), fun (37), humor (26), kind (25), active (17), outgoing (12), nice (11), easygoing (10)

**emotionality/eroticism/relationship (277):** life (32), friendship (29), loving (28), share (28), relationship (27), time (27), love (26), warm (20), romance (18), romantic (14), companionship (12), compassionate (8), long- (8)

**gender specificity (141):** man (56), gentleman (24), male (18), woman (18), lady (14), guy (11)

**adequacy/compatibility (345):** movies (41), interests (33), fit (31), music (29), travel (29), similar (23), art (18), theater (17), Jewish (15), smoker (15), 60s (12), children (10), sports (10), cultured (9), 40s (8), Italian (8), available (8), old (8), Asian (7), biking (7), concerts (7)

**education (0):** n/a **talent (0):** n/a

**vocation/financial assets (62):** professional (24), secure (22), successful (16)

**rationality/mental disposition (130):** intelligent (32), honest (25), sincere (18), down-to-earth (13), independent (13), smart (12), happy (9), witty (8)

### **6.3.3 All lexical items used by heterosexual men, grouped by semantic field**

**appearance (176):** attractive (52), slim (20), handsome (16), healthy (12), tall (12), brown (11), active (10), pretty (9), good- (8), blue (7), well- (7), 220lbs (6), complexion (6)

**social behavior/charisma (170):** good (42), fun (41), open (24), kind (12), easygoing (11), active (10), humor (9), quiet (8), nice (7), friends (6)

**emotionality/eroticism/relationship (311):** friendship (33), life (33), romance (25), open (24), time (23), love (22), relationship (21), romantic (18), loving (17), companionship (13), dates (13), times (13), affectionate (10), dinner (9), long (9), heart (8), special (7), term (7), compassionate (6)

**gender specificity (109):** woman (52), lady (29), female (14), man (14)

**adequacy/compatibility (249):** share (28), movies (25), music (25), fit (22), interests (20), travel (18), old (14), activities (13), conversation (13), similar (12), smoker (12), sports (12), age (9), same (9), things (9), 40s (8)

**education (0):** n/a **talent (0):** n/a

**vocation/financial assets (37):** secure (12), financially (10), professional (8), successful (7)

**rationality/mental disposition (93):** intelligent (23), honest (20), down-to-earth (16), healthy (12), easygoing (11), independent (11)

#### 6.4 Penn Treebank POS tagset (Marcus et al. 1993):

1. CC	Coordinating conjunction	25. TO	<i>to</i>
2. CD	Cardinal number	26. UH	Interjection
3. DT	Determiner	27. VB	Verb, base form
4. EX	Existential <i>there</i>	28. VBD	Verb, past tense
5. FW	Foreign word	29. VBG	Verb, gerund/present participle
6. IN	Preposition/subordinating conjunction	30. VBN	Verb, past participle
7. JJ	Adjective	31. VBP	Verb, non-3rd ps. sing. present
8. JJR	Adjective, comparative	32. VBZ	Verb, 3rd ps. sing. present
9. JJS	Adjective, superlative	33. WDT	<i>wh</i> -determiner
10. LS	List item marker	34. WP	<i>wh</i> -pronoun
11. MD	Modal	35. WP\$	Possessive <i>wh</i> -pronoun
12. NN	Noun, singular or mass	36. WRB	<i>wh</i> -adverb
13. NNS	Noun, plural	37. #	Pound sign
14. NNP	Proper noun, singular	38. \$	Dollar sign
15. NNPS	Proper noun, plural	39. .	Sentence-final punctuation
16. PDT	Predeterminer	40. ,	Comma
17. POS	Possessive ending	41. :	Colon, semi-colon
18. PRP	Personal pronoun	42. (	Left bracket character
19. PP\$	Possessive pronoun	43. )	Right bracket character
20. RB	Adverb	44. "	Straight double quote
21. RBR	Adverb, comparative	45. '	Left open single quote
22. RBS	Adverb, superlative	46. "	Left open double quote
23. RP	Particle	47. '	Right close single quote
24. SYM	Symbol (mathematical or scientific)	48. "	Right close double quote

#### 6.5 MATRI database

(Kirchhof, 2007), on DVD and <http://olive.spectrum.uni-bielefeld.de/tamino/MATRI>