Gen-X *So*

Douglas Kenter, Eric Lee, Rowyn McDonald
Linguistics 157
Professor Arnold Zwicky
March 23, 2007

1   Introduction

Although the use of *so* as an intensifier was first mentioned in linguistic literature in 1901 by Stoffel, its recorded use dates back over a thousand years to *Beowulf* (Tagliamonte and Roberts 2005). Throughout most of its history, *so* could be used as an intensifying adverb only to modify scalar adjectives, in such phrases as:

(1)   *You are so cool*.
(2)   *That is so awesome*.

Tagliamonte and Roberts (2005) note that many intensifiers, such as *really* or *pretty*, have often been considered nonstandard, and intensifier *so* is no exception, despite its long history of usage. However, intensifier use is prone to change over time, as certain words usurp others in popularity and acceptability. Tagliamonte and Roberts (ibid) have documented such a situation with *so*, which they concluded is replacing *really* as the most popular intensifier in North American speech. *Really* had previously replaced *very* (Tagliamonte and Roberts 2005).

In addition to its rise in frequency in colloquial usage, changes in who uses intensifier *so* and in what situations have also been observed. Along with other intensifiers, *so* has long been considered "feminine," although conclusive evidence that women actually use *so* more than men do is hard to come by (Kuha 2004, Pettibone 2004). In their longitudinal study of speech on *Friends*, Tagliamonte and Roberts (2005) saw a reduction of gender differences in the use of *so* over time. Furthermore, they saw it being used with more adjectives, an effect that is indicative of the process of delexicalizing the word. As a consequence of delexicalization, an intensifier can expand

the range of adjectives it modifies, gradually losing its lexical restrictions and becoming more frequent (Lorenz 2002).

Perhaps by extension of this phenomenon, a new construction has emerged in the past two decades, termed "Gen-X *So*" by Arnold Zwicky.[1] Gen-X *So*, which is so named because of its initial appearance in the speech of members of "Generation X," a popularize term for the generation of Americans born roughly between the years of 1963 and 1978.[2] Gen-X *So* greatly extends the use of intensifier *so* and allows it to occur in a number of grammatical environments where use of *so* is traditionally considered nonstandard (Pettibone 2004). The online Oxford English Dictionary acknowledges this recent development in its draft changes from 2006, which include the addition below for the *so* entry:

> *slang* (chiefly US). As an intensifier, forming non-standard grammatical constructions.[3]

The earliest recorded example is from the 1988 film *Heathers* ("Grow up, Heather. Bulimia's so '86"), which has been credited with facilitating its spread (Zwicky 2006).

Pettibone (2004) outlines the new structural environments in which Gen-X *So* is observed to occur. The range of structures described in that paper, along with illustrative and invented examples, are listed here:

(3)   *so* + non-scalar adjectives
            You are so dead.

(4)   *so* + nominal phrase complement

[1] http://listserv.linguistlist.org/cgi-bin/wa?A2=ind0103E&L=ADS-L&P=R3043&I=-3
[2] http://www.tedrall.com/longarticle_004.htm
[3]
http://dictionary.oed.com/cgi/entry/50229610?query_type=word&queryword=so&first=1&max_to_show=10&sort_type=alpha&result_place=2&search_id=0KkQ-2yAHQn-357&hilite=50229610

That is so last year.

(5)   *so* + prepositional phrase complement

I am so out of the loop.

(6)   *so* + verb phrase complement

I am so getting that.

I so rock at this.

For canonical *so*, these constructions would have all been considered colloquial and nonstandard. However, like canonical *so*, Gen-X *So* can modify scalar adjectives. The use of Gen-X *So* in these cases is only distinguishable from canonical *so* by the distinctive intonation of Gen-X *So* and by the structure of its negation. In all of these structures, negation is achieved by placing not immediately after *so*, which is not usually the case for canonical *so* when modifying scalar adjectives.

How these novel uses of *so* arose has not been definitively determined, but Zwicky (2000) hypothesizes that a reanalysis of intensifier *so* took place. When *so* is used to modify an adjective, it can be understood as a modifier of the adjective within the adjective phrase, or it can be reanalyzed as a modifier of the adjective phrase itself, thus moving up a node in the syntactic structure. From there, *so* could be extended to modify other types of predicative phrases, or it can extend the types of verbs it can co-occur with (Zwicky 2000, Pettibone 2004). This type of process fits the delexicalization trajectory described by Tagliamonte and Roberts (2005), where an intensifier is initially limited to having certain collocations but gradually becomes less lexically restricted.

Both Kuha (2004) and Pettibone (2004) comment on the position of Gen-X *So* with respect to the verb. While Gen-X *So* can occur before or after a copula *be* or an

auxiliary, it can only appear before the verb in all other cases. So, while (7) and (8) are

considered acceptable by some speakers,

(7)   *That is so funny!*

(8)   *That so is funny!*

between (9) and (10), only the first structure is ever attested.

(9)   *I so need to go home.*

(10)   **I need so to go home.*

Pettibone (2004) explores the acceptability of the different possible structures using Gen-

X *So* in some detail. The study compared how the acceptability judgments of having so

before or after the verb depended on the type of verb used. The acceptability judgments

of the sentences differed depending on which age group the person was in.

One of the questions raised by previous work on Gen-X *So* is whether it can be

used to modify different types of predicative complements with equal acceptability. It

can be predicted that, given the gradual nature of an intensifier's delexicalization, the

newer Gen-X *So* structures would be considered less acceptable than ones that have been

around longer. We will explore these dynamics of acceptability between different types

of predicative complements used with Gen-X *So*.

This paper also takes a longitudinal look at the use of Gen-X *So* in television

shows and movies over a period of roughly twenty years, from the late 1980s until the

present day. Depictions of language usage in the media can be a good indicator of what

is currently happening in everyday speech (Tagliamonte and Roberts 2005). Which

social groups are popularly associated with Gen-X So can therefore be ascertained by

observing the types of speakers that use it in the media.

2   Actual usage

Lacking the resources to collect traditional sociolinguistic data, we turned to media language as a more readily available source of data. Loosely following the precedent of Tagliamonte and Roberts (2005), who argue that "television data can provide interesting and informative sociolinguistic data" (280), a small corpus of Gen-X *So* uses was collected from American movies and television shows dating from 1991 to the present. These examples come from unofficial transcripts of movies and television shows available on the Internet[4]. In order to restrict the examples to clear cases of Gen-X *So*, the following search terms were used:

(11)   *so not* (e.g. "*That is so not cool*")

(12)   *I/he/she/we/they so* (e.g. "*I so never, ever promised*")

In a more thorough study, it would be preferable to systematically sift through each transcript to extract all occurrences of *so*; the current study's method resulted in a limited set of data which was dominated by examples of *so* modifying verbs. A total of 114 occurrences were collected from 92 different shows and movies. Each token was tagged for the type of phrase *so* modified, the year of release, and speaker age, gender, and ethnicity (to the extent that this information was available). The potential interest of the media data is twofold. First, portrayal of *so* users in the media may provide insight into popular perceptions of the word. Second, the use of *so* in film and media may facilitate its spread, as has been suggested with regard to the film *Heathers*, which contains one of the earliest recorded uses of Gen-X *So*.

---

[4] The transcripts were found primarily on http://twiztv.com/scripts/, http://www.script-o-rama.com, and http://www.dailyscript.com

So, like intensifiers in general, is often strongly associated with young, female speakers. Thus we predicted that our data would reflect this perception and include more examples from female than male speakers, and more examples from younger (teens and early twenty-somethings) than older speakers. While our results are merely suggestive rather than conclusive, we did find that these some of these predictions held. Gender data was only available for 58 tokens of *so*, but of these around 75% were spoken by female speakers, which is statistically different from the prediction that men and women use so equally with a chi-squared value less than 0.01. The gender results are shown in table 1.

TABLE 1

	Number	Percentage
Female	44	**75.86%**
Male	14	24.14%
Total	58	100.00%

Data on speaker age was more difficult to obtain; only 28 tokens were tagged. No tokens were recorded for speakers below the age of 10 or above the age of 35, and the remaining tokens were broken down into the following age brackets: 10-14, 15-19, 20-24, 25-30, and 30-35.

As with gender, the data is suggestive The most tokens, 32.1%, come from speakers age 15 to 19, followed by 21.4% from speakers age 20 to 24. Together these two groups account for over half the tokens, which seems to support the hypothesis that Gen-X *So* is more prevalent among younger speakers. Ideally it would be best to normalize this data against the age distribution of (speaking) characters from the shows and movies surveyed, but as such demographic data is hard to come by they will instead be compared to data from the 2000 United States Census. The census does not have

separate statistics for ages 25 to 30 and 30 to 34, so these bins had to be combined into a single bin. Only the population from age 10 to 34 was considered, and the predicted values shown in figure 1 are based on the percentage that each age group constitutes of this restricted population.

FIGURE 1
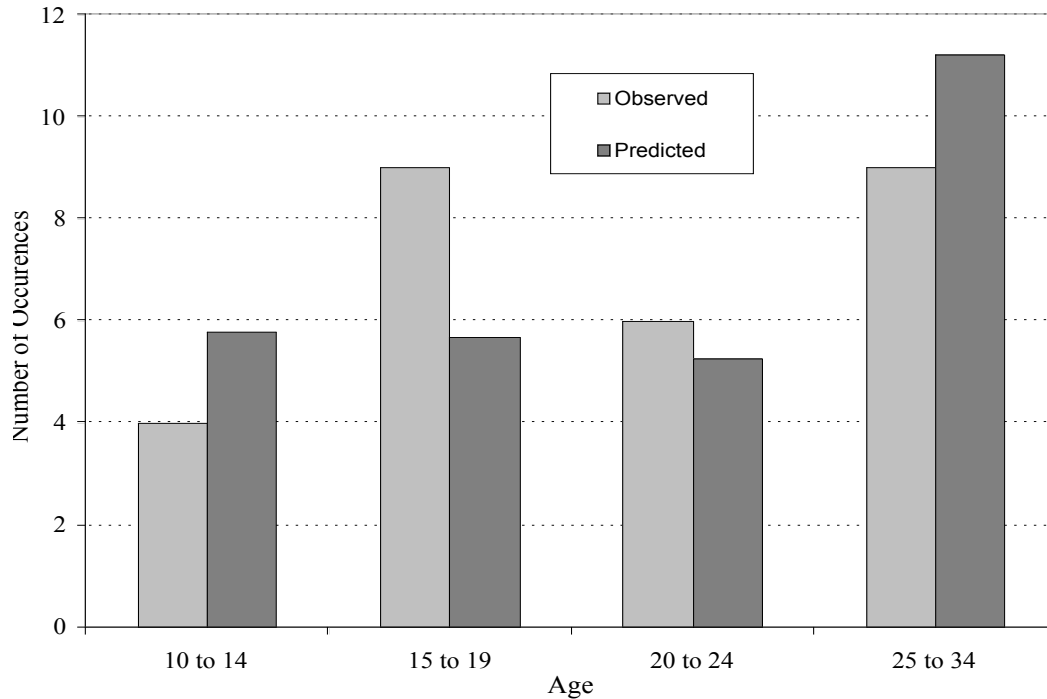
The chi-squared significance level was 0.388, indicating that the distribution of *so* usage by age in the data was not statistically significant.

The final sociolinguistic variable considered in this study was ethnicity. No previous studies regarding *so* mention ethnicity, so we made predictions based on our own intuitions. We predicted that Gen-X *So* speakers would be primarily white, and that Gen-X *So* would not be commonly used by speakers of other ethnicities. Again the raw data appears to substantiate this claim; of 39 tokens tagged for ethnicity, 37 (94.87%) were white, 1 was black or African American, and 1 was Asian. To provide perspective, these results were compared to data from the 2000 US Census. The observed values

differed from the values predicted by the census with a chi-squared value of 0.076, which implies that ethnicity was in fact not a significant factor in predicting Gen-X *So* usage.

In summary of the sociolinguistic results, the only factor found to be statistically significant in predicting the use of Gen-X *So* was gender. In holding with common perception, *so* was used more frequently by females than males. This may not be an effect specific to *so*, however, because intensifier use in general is associated with women, and we did not compare *so* to other intensifiers from the same data pool in this study. It is also worth noting that the results might be different for any of the three sociolinguistic variables we considered if they were compared to demographic data specific to the media from which the tokens were collected. Instead, we assumed that the demographics of characters across a large sample of media (nearly 100 films and shows) would be roughly the same as the general population demographics of the United States, and use the census data as a basis for comparison.

In their study of intensifiers in the television series *Friends*, Tagliamonte and Roberts (2005) found that the use of *so* followed an unusual trajectory between across the years 1994 to 2002. *So* generally rose in frequency until 1998, but then dipped in 1999 and 2000 before regaining popularity in 2001. They correlate this trend with the popularity of *Friends*, which declined at the same time as the frequency of *so*. In our data, the distribution of *so* across the years 1991 to 2007 was also unusual but did not match that of Tagliamonte and Roberts. Figure 2 shows the number of occurrences for each of the years between 1991 and 2007.

FIGURE 2
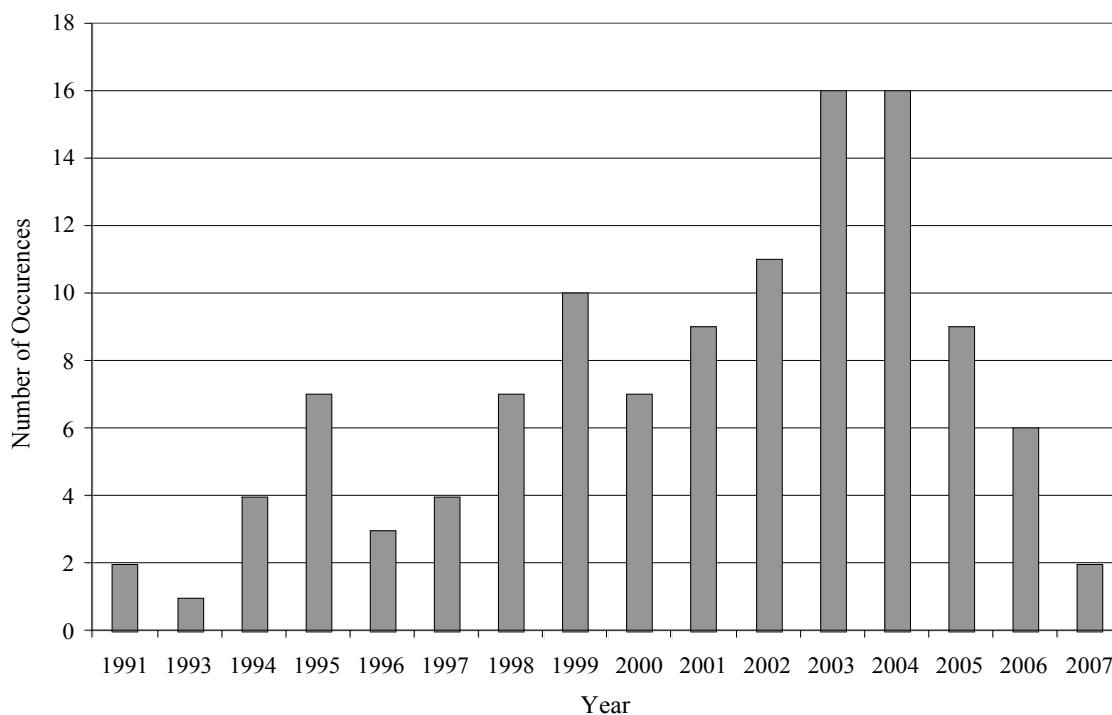
Interestingly, *so*'s frequency appears to peak in 2003-2004 before declining sharply over the next three years, to the present day. We were unable to generate any linguistic hypotheses for this trend; instead, it may be affected by several non-linguistic factors, the most likely of which is the availability of the online transcripts from which the data were collected. Because the transcripts are unofficial and usually written by fans of the movies and television shows, it would make sense if there were fewer transcripts available for more recent releases, which would help explain the apparent decline of *so* in the past two years.

3   Perceived Usage

To elicit perceptions of Gen-X *So* usage, we created a survey (See Appendix A), which requested judgments of the acceptability of Gen-X *So* in different syntactic positions: preceding a verb phrase (VP) predicative complement, preceding a noun phrase (NP) predicative complement, preceding an adjectival phrase (AP) predicative

complement, and preceding a prepositional phrase (PP) predicative complement. We

also asked for a subjective judgment of the hypothetical *so* speaker's gender, sexuality,

urban or rural status, age, and ethnicity.

The results of the survey are both intuitive and surprising. The extremely high

association of Gen-X *So*, in all syntactic environments, with youth, white ethnicity and

urban status (97% youth association, 94% white ethnic association, nearly the same as the

rate , 98% urban association), aligns with its stereotypical use among young urban

speakers. We also noted high associations with women (76%, again, the same percentage

as we discovered from the media study much like the association with white ethnicity)

and heterosexuality (87%). Though some of these categories are commonly perceived as

discrete and binary (male/female, gay/straight) and others more fluid (urban/rural,

young/old, white/non-white), we elected to let the subject make his or her own judgment

about the thresholds between each category pair. The combined percentages between

category pairs may go above 100%, as subjects were allowed to mark both options.

Gen-X *So* Associated Categories	Percent of Subjects Listing Association
Male	41%
Female	76%
Gay	9%
Straight	87%
Urban	98%
Rural	4%
Young	97%
Old	5%
White	94%
Non-white	5%

With 29 subjects evaluating 9 different Gen-X So constructions, 251 judgments of

acceptability were made, on a scale of 1 to 5, with a score of 1 being completely

acceptable grammatically and a score of 5 being completely unacceptable grammatically.

The average grammatical acceptability of Gen-X *So*, across all syntactic categories, was

2.37, slightly under the 2.5 of a "neutral" or "average" acceptability, placing Gen-X *So* in

acceptable grammatical territory, on average — though just barely. Within Gen-X *So*

syntactic categories, the most acceptable phrases employing Gen-X *So* had a VP

predicative complement, followed by AP predicative complements, followed by PP

predicative complements, with NP predicative complements considered least acceptable.

Gen-X So Predicative Complement	Average Acceptability Score
Verb Phrase	*2.20*
Adjectival Phrase	*2.28*
Prepositional Phrase	*2.48*
Noun Phrase	*2.71*

Gen-X *So* Sentence (type predicative complement)	Average Acceptability Score
I am so going to hell for this link. (VP)	*2.00*
I am so not kidding about this. (VP)	*2.00*
You are so dead. (AP)	*2.14*
You are so last year. (NP)	*2.24*
I am so not ready for this. (AP)	*2.31*
I am so ridiculously broke in S.F. right now. (AP)	*2.38*
You are so off my buddy list. (PP)	*2.48*
I so hate men right now. (VP)	*2.59*
I am so a dinosaur. (NP)	*3.17*

However, though *so* with an NP predicative complement was, on average,

considered least acceptable by the subjects among all the *so* possibilities offered, the two

so sentences followed by an NP predicative complement on the survey were quite

different from one another in acceptability, which allowed us to frame a compelling sort

of "minimal semantic pair" of phrases of the syntactic construction of Gen-X *So*

preceding an NP predicative complement. In other words, the acceptability score for *so*

followed by an NP predicative complement can be explained by looking at the component sentences, which have different acceptability for different reasons.

(13)   *I am so a dinosaur.*

(14)   *You are so last year.*

First, sentence, (13) is of vague semantic character, and is an NP fronted by a determiner, two characteristics of the construction that make it less acceptable than sentence (14), and ungainly in the ears of many subjects. It becomes more salient, however, when it is cast against sentence (14), which, as a sentential vessel for Gen-X *So*, is not as novel or innovative as sentence (13), and with good reason: it is closer to an "idiom unit" than an organically-constructed syntactic structure. In the fashion of "Grow up Heather, bulimia's so '86," and perhaps due to its similarity to this kind of "core," primeval use of Gen-X *so*, sentence (14) was perhaps considered by the subjects to be a singular syntactic chunk, if only in the semantic domain of referring to a time in the past to "date" an object or person. In other words, Gen-X *so* seems to be making inroads to linguistic acceptability through its original innovative phrases which have become acceptable as they become "canon" among idioms. Also, VP predicative complements, which may also derive their acceptability from being some of the first syntactic configurations of Gen-X *So*, seem to be at the vanguard of acceptable types of Gen-X *So* constructions.

4   Conclusion

Gen-X *So*, though only about twenty years old, has had a robust life, and appears to be gaining acceptability in a variety of syntactic configurations, while remaining less acceptable in others. It will most likely remain colloquial for the next few generations, as

stereotyped associations die hard, especially those with near-100% affinity with youth

and urbanity. It is also worth repeating that intensifier *so*, though widely accepted, is still

considered colloquial. We expect Gen-X *So* to slowly become assimilated into the

lexicon until it is eventually bleached of any overt gender, age, regional or sexuality

associations, at which point it will also have equal syntactic fluency with predicative

complements, whether NP, AP, PP or VP. Slightly contradicting the stereotype of near-

total use among females, Gen-X *So*'s gender association with appears to be weaker than

its associations with age, ethnicity and region, evidenced by data in both the media study

and our survey. Still, the erratic frequency behavior of Gen-X *So*, in spite of its general

trend of increasing in frequency, underscore its nascence as an innovation, and remind us

that its possible eventual acceptance and use cannot be firmly predicted yet.

References

Kuha, Mai. 2004. Investigating the Spread of "so" as an Intensifier: Social and Structural
Factors. Presented at the Symposium about Language and Society – Austin
(SALSA) XXI, April 16-18.

Pettibone, Jeanette. 2004. Gen-X 'so.'

Tagliamonte, Sali and Chris Roberts. 2005. So Weird; So Cool; So Innovative: The Use
of Intensifiers in the Television Series *Friends*. *American Speech*, Vol. 80, No. 3.

Zwicky, Arnold. 2000. "Re: Sheidlower in NY Times." Posted August 31, 2000 to ADS-
L: http://listserv.linguistlist.org/cgi-bin/wa?A2=ind0008E&L=ads-l&P=R1863

Zwicky, Arnold. 2006. "So In Style at the NYT." Posted April 6, 2006 to Language Log:
http://itre.cis.upenn.edu/~myl/languagelog/archives/002994.html

Appendix A: Gen-X *So* Survey

Age ___

Gender ___

Ethnicity ___

Home Region ___

Consider the following nine sentences. Please write how acceptable you think the sentence is on the scale provided, and write what kind of person, and from where, you think would say such a phrase. Mark all options that apply.

1   *I am so going to hell for this link.*

   Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____

2   *I so hate men right now.*

   Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____

3   *I am so ridiculously broke in S.F. right now.*

   Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___ Rural ___   Young ___   Old ___

Ethnicity _____

4   *You are so dead.*

    Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___ Rural ___   Young ___   Old ___

Ethnicity _____

5   *You are so off my buddy list.*

    Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___ Rural ___   Young ___   Old ___

Ethnicity _____

6   *I am so a dinosaur.*

    Totally Acceptable   1   2   3   4   5   Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____

7   *You are so last year.*

 Totally Acceptable 1 2 3 4 5 Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____

8   *I am so not ready for this.*

 Totally Acceptable 1 2 3 4 5 Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____

9   *I am so not kidding about this.*

 Totally Acceptable 1 2 3 4 5 Totally Unacceptable

Perceived Region _____

You would expect a person who says the above sentence to be:

Male ___   Female ___   Gay ___   Straight ___

Urban ___   Rural ___   Young ___   Old ___

Ethnicity _____