Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation

Dan Jurafsky
Linguistics Department
Stanford University
jurafsky@stanford.edu

Rajesh Ranganath
Computer Science Department
Stanford University
rajeshr@stanford.edu

Dan McFarland
School of Education
Stanford University
dmcfarla@stanford.edu

Abstract

Automatically extracting social meaning and intention from spoken dialogue is an important task for dialogue and information extraction applications. We describe a system for solving the new task of detecting elements of **interactional style**: whether a speaker is *awkward*, *friendly*, or *flirtatious*. We create and use a new spoken corpus of approximately 1000 4-minute speed-dates. Participants rated their interlocutors for these elements of style. Despite the difficulty of automatically determining interactional style from noisy natural data, we are able to detect flirtatious, awkward, and friendly styles in conversations with between 55% and 65% accuracy, compared to a 50% baseline, based on prosodic, lexical, and discourse features. We describe simple ways to extract relatively rich dialogue features, and analyze which features performed similarly for men and women and which were gender-specific.

1 Introduction

How can we extract social meaning from speech, deciding if a speaker is particularly engaged in the conversation, is uncomfortable or awkward, or is particularly friendly and flirtatious? Understanding these meanings and how they can be extracted from conversation is an important task in itself, and is also key toward understanding how we can build more natural and more convincing dialog agents.

Cues for social meaning permeate speech at every level of linguistic structure, and lexical, prosodic, and discourse-structural information can all be used as cues for many different kinds of social meaning. For example, acoustic cues such as low and high F0 or energy and spectral tilt are important in detecting emotions such as *annoyance*, *anger*, *sadness*, or *boredom* (Ang et al., 2002; Lee and Narayanan, 2002; Liscombe et al., 2003) as well as speaker characteristics such as the presence of *charisma* (Rosenberg and Hirschberg, 2005). Lexical cues to social meaning abound. Speakers with links to depression or speakers who are under stress use more first person singular pronouns (Rude et al., 2004; Pennebaker and Lay, 2002; Cohn et al., 2004), while negative emotion words are useful cues to deceptive speech (Newman et al., 2003). Finally, other dialog features such as the presence of disfluencies can inform listeners about speakers' problems in utterance planning or about confidence (Brennan and Williams, 1995; Brennan and Schober, 2001).

Our goal in this project is to see whether cues of this sort are successful in detecting particular elements of conversational style and social intention; whether a speaker in a speed-dating conversation is judged by the interlocutor as *friendly*, *awkward*, or *flirtatious*.

2 The Corpus

Our experiments make use of a new corpus we have collected, the SpeedDate Corpus, collected and annotated over the last 3 years. The corpus is based on three speed-dating sessions run at an elite private American university in 2005. The graduate student participants volunteered to be in the study and

were promised emails of persons with whom they reported mutual liking. Each date was conducted in an open setting where there was substantial background noise. All participants wore audio recorders on a shoulder sash, thus resulting in two audio recordings of the approximately 1000 4-minute dates. In addition to the audio, we collected pre-test surveys, event scorecards, and post-test surveys. This is the largest sample we know of where audio data and detailed survey information were combined in a natural experiment.

The survey information gives us a rich collection of information, including about the success of the conversation (whether participants thought the conversation 'clicked', their enjoyment of the date, whether they had shared interests, and who was interested in dating each other) but also includes labels from each speaker on the conversational style and intention of the interlocutor (awkward, friendly, flirtatious, funny, and assertive). Each speaker was asked to report how often their dates speech reflected different conversational styles on a scale of 1-10 (1=never, 10=constantly): "How often did the other person behave in the following ways on this 'date'?". A variety of other items were also included concerning the respondents perceptions of their dates attractiveness, intelligence, sincerity, courteousness, ambition, and humor. Pretest and posttest surveys concerned general attitudes about dating, preferences in mates, and demographic information.

Identification of conversational styles entailed linking the aforementioned survey data to acoustic information. We acquired acoustic information by taking the acoustic wave file from each recorder and manually segmenting it into a sequence of wavefiles, each corresponding to one 4-minute date. Since both speakers wore microphones, most dates had two recordings, one from the male recorder and one from the female recorder. Because of mechanical, operator, and experimenter error, some of the recordings were lost, and thus some of the dates had only one recording. Transcribers working for a professional transcription service used the two recordings to create a transcript for each date, and time-stamped the start and end time of each speaker turn. Transcribers were instructed to mark various disfluencies as well as some non-verbal elements of the conversation such as laughter.

Because of noise, malfunctioning recorders, and some segmentation and transcription errors, a number of dates were not possible to analyze. The current study focuses on the 992 remaining clean dates for which we had usable audio, transcripts, and survey information.

3 The Experiments

Our goal is to detect three of the style variables, in particular awkward, friendly, or flirtatious speakers, via a machine learning classifier. Recall that each speaker in a date (each conversation side) was labeled by his or her interlocutor with a rating from 1-10 for awkward, friendly, or flirtatious behavior. For the experiments, the 1-10 Likert scale ratings were first mean-centered within each respondent so that the average was 0. Then the top ten percent of the respondent-centered meaned Likert ratings were marked as positive for the trait, and the bottom ten percent were marked as negative for a trait. Thus each respondent labels the other speaker as either positive, negative, or NA for each of the three traits

We run our binary classification experiments to predict this output variable.

For each speaker side of each 4-minute conversation, we extracted features from the wavefiles and the transcript, as described in the next section. We then trained three separate binary classifiers for the 3 tasks, as described in Section 5.

4 Feature Extraction

In selecting features we drew on previous research on the use of relatively simple surface features that cue social meaning, described in the next section.

Each date was represented by the two 4-minute wavefiles, one from the recorder worn by each speaker, and a single transcription. Because of the very high level of noise, the speaker wearing the recorder was much clearer on his/her own recording, and so we extracted the acoustic features for each speaker from their own microphone. All lexical and discourse features were extracted from the transcripts.

All features describe the speaker of the conversation side being labeled for style. The features for a conversation side thus indicate whether a speaker

who talks a lot, laughs, is more disfluent, has higher F0, etc., is more or less likely to be considered flirtatious, friendly, or awkward by the interlocutor. (All features were also computed for the interlocutor as well as we will discuss below.)

4.1 Prosodic Features

F0 and RMS amplitude features were extracted using Praat scripts (Boersma and Weenink, 2005). Since the start and end of each turn were time-marked by hand, each feature was easily extracted over a turn, and then averages and standard deviations were taken over the turns in an entire conversation side. Thus the feature F0 MIN is the average over all turns in the conversation side of the (non-zero) F0 min of each turn from the side. F0 MIN STD is the standard deviation across turns of this same measure.

F0 MIN	minimum (non-zero) F0 per turn, averaged over turns
F0 MIN STD	standard deviation from F0 min
F0 MAX	maximum F0 per turn, averaged over turns
F0 MAX STD	standard deviation from F0 max
F0 MEAN	mean F0 per turn, averaged over turns
F0 MEAN STD	standard deviation (across turns) from F0 mean
F0 STD	standard deviation (within a turn) from F0 mean, averaged over turns
F0 STD STD	standard deviation from the f0 std
PITCH RANGE	f0 max - f0 min per turn, averaged over turns
PITCH RANGE STD	standard deviation from mean pitch range
RMS MIN	minimum amplitude per turn, averaged over turns
RMS MIN STD	standard deviation from RMS min
RMS MAX	maximum amplitude per turn, averaged over turns
RMS MAX STD	standard deviation from RMS max
RMS MEAN	mean amplitude per turn, averaged over turns
RMS MEAN STD	standard deviation from RMS mean
TURN DUR	duration of turn in seconds, averaged over turns

Table 1: Prosodic features for each conversation side, extracted using Praat from the hand-segmented turns of each side.

Note that we coded four measures of f0 variation, not knowing in advance which one was likely to be the most useful: F0 MEAN STD is the deviation across turns from the global F0 mean for the conversation side, measuring how variable the speakers mean f0 is across turns. F0 STD is the standard deviation within a turn for the f0 mean, and then averaged over turns, hence measures how variable the speakers f0 is within a turn. F0 STD STD measures how much the within-turn f0 variance varies from turn to turn, and hence is another measure of cross-turn f0 variation. PITCH RANGE STD measures how much the speakers pitch range varies from turn to turn, and hence is another measure of cross-turn f0 variation.

4.2 Lexical Features

Only a small number of shallow lexical features were examined, as summarized in Table 2.

TOTAL WORDS	total number of words
YOU	uses of the word *you* not counting *you know*
WE	uses of the word *we*
PAST TENSE	uses of past tense auxiliaries *was, were, had*

Table 2: Lexical features; each is a total count for each conversation side.

4.3 Discourse Features

A number of discourse features were extracted, drawing from the conversation analysis, disfluency and dialog act literature (Sacks et al., 1974; Jurafsky, 2001). While discourse features are clearly important for extracting social meaning, previous work on social meaning has met with less success in use of such features (with the important exception of (Enos et al., 2007)), presumably because discourse features are expensive to hand-label and hard to automatically extract. We chose a number of discourse features that we felt could be automatically extracted but might give useful cues.

Three particular dialog acts were chosen as shown in Table 3. **Backchannels** (also known as *continuers*) and **appreciations** (a particular kind of continuer dialog act that express positive affect) were coded by hand-built regular expressions. The regular expressions were based on analysis of the

BACKCHANNELS	number of backchannel utterances in side (*Uh-huh.*, *Yeah.*, *Right.*, *Oh, okay.*)
APPRECIATIONS	number of appreciations in side (*Wow, That's true, Oh, great*)
QUESTIONS	number of questions in side
LAUGH	number of instances of laughter in side
TURNS	total number of turns in side

Table 3: Dialog act features.

backchannels and appreciations in the hand-labeled Switchboard corpus of dialog acts (Jurafsky et al., 1997). Questions were coded simply by the presence of question marks. We also included the number of instances of laughter for the side, as well as the total number of turns a speaker took.

A second group of discourse features relating to repair, disfluency, and related issues like speaker overlap are summarized in Table 4. Filled pauses

UH/UM	total number of filled pauses (*uh* or *um*) in conversation side
RESTART	total number of disfluent restarts in conversation side
OVERLAP	number of turns in side which the two speakers overlapped
NTRI	Next Turn Repair Indicator (*Wait*, *Excuse me*)

Table 4: Disfluency features

(*um*, *uh*) were coded by regular expressions (the transcribers had been instructed to transcribe all filled pauses). Restarts are a type of repair in which speakers begin a phrase, break off, and then restart the syntactic phrase. The following example shows a restart; the speaker starts a sentence *Uh, I* and then restarts, *There's a group of us...*:

> Uh, I–there's a–a group of us that came in the same year–

Overlaps are cases in which both speakers were talking at the same time, and were marked by the transcribers in the transcripts:

MALE:	But-and also obviously–
FEMALE:	It sounds bigger.
MALE:	–people in the CS school are not quite as social in general as other–

Finally, next turn repair indicators (NTRIs) are turns in which a speaker signals lack of hearing or understanding (Schegloff et al., 1977). Since we did not have full dialog act labels, we used a simple approximation to these, consisting of speakers using 'Excuse me' or 'Wait', as in the following examples:

FEMALE:	Okay. Are you excited about that?
MALE:	Excuse me?

MALE:	And then they–oh yeah.
FEMALE:	Wait say it again, do you have another one?

5 Classifier Training

We used regularized logistic regression for our classification.

Before performing the classification task, we preprocessed the data in two ways. First, we standardized all the variables to have zero mean and unit variance. We did this to avoid imposing a prior on any of the features based on their numerical values. Consider a feature A with mean 100 and a feature B with mean .1 where A and B are correlated with the output. Since regularization favors small weights there is a bias to put weight on feature A because intuitively the weight on feature B would need to be 1000 times larger to carry the same effect. Second, we removed highly correlated features, One goal of removing correlated features was to remove as much colinearity as possible from the regression so that the regression weights could be analyzed to rank features for their importance in the classification. In addition, we hoped to improve classification since the number of training examples required to train well with more features is larger (Ng, 2004). Empirically, removing correlated features also seemed to produce better overall performance, as we discuss below. We accomplished correlated feature removal by looking at all pairs of features and rejecting one of the features in the pair if the correlation is greater than .7.

For each classification task due to the small amounts of data we performed k-fold cross validation to learn and evaluate our models. We used a variant of k-fold cross validation with five folds where three folds are used for training, one fold is used for validation, and one fold is used as a test set. This test fold is not used in any training step. The yields a datasplit of 60% for training, 20% for validation, and 20% for testing. This yields 120 training examples, 40 validation examples, and 40 test examples.

Recall that in logistic regression we train a vector of feature weights $\theta \in \mathbb{R}^n$ so as to make the following classification of some output variable y for an input observation x:

$$p(y|x;\theta) = \frac{1}{1+\exp(-\theta^T x)} \quad (1)$$

In regularized logistic regression we find the weights θ which maximize the following optimization problem:

$$\underset{\theta}{\operatorname{argmax}} \sum_i \log p(y^i|x^i;\theta) - \alpha * R(\theta) \quad (2)$$

$R(\theta)$ is a regularization term used to penalize large weights. We chose $R(\theta)$, the regularization function, to be the L_1 norm of θ. That is, $R(\theta) = ||\theta||_1 = \sum_{i=1}^{n} |\theta_i|$.

In our case, given the training set S_{train}, test set S_{test}, and validation set S_{val}, we trained the weights θ as follows:

$$\underset{\alpha}{\operatorname{argmax}} \operatorname{accuracy}(\theta_\alpha, S_{\text{val}}) \quad (3)$$

where for a given sparsity parameter α

$$\theta_\alpha = \underset{\theta}{\operatorname{argmax}} \sum_i \log p(y^i|x^i;\theta) - \alpha * R(\theta) \quad (4)$$

We chose L_1-regularization because the number of training examples to learn well grows logarithmically with the number of input variables (Ng, 2004), and to achieve a sparse activation of our features to find only the most salient explanatory variables. This choice of regularization was made to avoid the problems that often plague supervised learning in situations with large number of features but only a small number of examples. The search space over the sparsity parameter α is bounded around an expected sparsity to prevent overfitting.

Finally, to evaluate our model on the learned α and θ_α we used the features X of the test set S_{test} to compute the predicted outputs Y using the logistic regression model. Accuracy is simply computed as the percent of correct predictions.

6 Results

Results for the 6 binary classifiers are presented in Table 5.

	Awk		Flirt		Friendly	
	M	F	M	F	M	F
Speaker	65%	56%	55%	61%	63%	59%
+other	56%	55%	52%	63%	66%	55%

Table 5: Accuracy of binary classification of each conversation side, where chance is 50%. The first row uses features only from the single speaker; the second adds all the features from the interlocutor as well.

The first row shows results using features extracted from the speaker being labeled. In general, awkwardness is easiest to detect in men, flirtatiousness in women, and friendliness in men.

The second row of table 5 shows the accuracy when using features from both speakers. Surprisingly, adding information about the interlocutor tends to make classification worse. In fact, interlocutor features only improve the detection of friendliness in men and flirtatiousness in women. It's not clear whether this is because the interlocutor linguistic behavior didn't generalize well to the test set, or whether the additional, possibly collinear variables in general caused overfitting problems for the classifier; in general adding more features requires more training data to successfully train a model.

We next considered the features that helped classification. Table 6 shows feature weights for the features (features were normed so weights are comparable), and is summarized in the following paragraphs:

- Men who are labeled as awkward have a narrower, monotonous intensity range (they don't get nearly as quiet or as loud), laugh less, take fewer turns of longer duration with less words (hence presumably slower rate of speech), use

MALE AWK		MALE FLIRT		MALE FRIENDLY	
RMS min	0.1653	we	0.0549	laugh	0.1204
appreciation	-0.1359	RMS min std	-0.0513	overlaps	0.1021
f0 min	-0.0541	f0 mean	-0.0486	RMS min std	0.0646
we	0.0457	laugh	0.0359	backchannel	-0.0502
turns	-0.0401	overlaps	-0.025	f0 min	0.0437
question	-0.0393	f0 std std	0.0253	we	-0.0393
RMS max	-0.0336	question	0.0214	f0 std	-0.037
past	-0.0324	backchannel	-0.0209	past	0.0288
laugh	-0.0228	past	0.0203	turn dur	0.0223
uh	0.0194	RMS max	-0.0166		
f0 mean	-0.0149	pitch range std/f0 max std	0.0029		
turn dur	0.0107				
NTRI	-0.0099				

FEMALE AWK		FEMALE FLIRT		FEMALE FRIENDLY	
backchannel	-0.1132	laugh	0.187	laugh	0.1363
f0 max/pitch range	-0.0903	RMS min	-0.1218	backchannel	0.0629
turn dur	0.0592	uh	-0.1009	RMS min	0.0569
wordcount	-0.0517	past	0.0807	uh	-0.0483
overlap	-0.0438	f0 std	-0.0807	turn dur	-0.0468
NTRI	-0.0408	NTRI	0.0617	RMS max	-0.0355
uh	0.0401	turns	-0.0546	pitch std std	0.0307
question	0.0239	appreciation	-0.0482	overlaps	-0.0126
past	0.0238	f0 mean	0.047	RMS min std	0.0104
appreciation	0.0155	we	-0.0401	wordcount	-0.0057
we	-0.0148	backchannel	0.0358	past	-0.0007
f0 std std	-0.0138	question	0.0368		
f0 mean	-0.007	overlaps	0.0091		
RMS min std	-0.0023	you	0.0034		

Table 6: Feature weights for the most significant predictors for each classifier.

more *uh*s, don't ask questions, don't use appreciation backchannels or NTRIs, and have lower pitch.

- Women who are labeled as awkward, like awkward men, are likely to use filled pauses (*uh* and *um*), and take longer turns that have fewer words (suggesting that they are speaking more slowly), and use lower pitch. Unlike awkward men, awkward women ask more questions, and they don't backchannel. Where awkward men lower their min f0, awkward women lower their max f0 or pitch range

- Men who are labeled as flirting laugh more, ask more questions, use "we", use the past tense, are less likely to overlap, don't use simple backchannels, have a lower mean pitch and don't get quite as loud. They also have a more variable pitch, by two measure (increase in pitch range and greater cross-turn variation of their within-turn pitch variation).

- Women who are labeled as flirting, like men, laugh more, ask more questions, and use the past tense more. Unlike men, they are more likely to overlap and they do use simple backchannels. Also unlike men, women labeled as flirting have a higher average f0 and a much lower amplitude minimum (that is, they get much softer).

- Men who are labeled as friendly laugh more, take longer turns, overlap more, use the past, don't use simple backchannels, and have higher minimum pitch (i.e., don't go as low).

- Women labeled as friendly laugh more, but are otherwise quite different than men who are labeled as friendly. Women are less likely to overlap, more likely to use simple backchannels, have a reduced intensity range (don't get as soft or as loud), and take shorter turns with less words.

7 Discussion

The results presented here should be regarded with some caution. The sample is not a random sample of English speakers or American adults. As such, further collection and analysis is probably needed. In addition, speed dating is not a natural context for expressing and witnessing every conversational style and social intent. Therefore, a wider array of recordings across settings and encounters would be required before a more general theory of conversational styles is established.

On the other hand, the presented results may under-reflect the relations being captured. The quality of recordings and coarse-grained time-stamps likely cloud the relations, and as the data is cleaned and improved, we expect the associations to only grow stronger. Moreover, it is feasible to believe future renditions of this study could be performed in a more automated means to overcome quality concerns and fill in perceived data / sampling gaps.

Caveats aside, we believe the evidence indicates that the perception of several types of conversational style have relatively clear signals across genders, but with some additional gender contextualization.

For both genders, awkwardness is conveyed through slow, lower pitched, stilted talk. Similarly, both genders convey flirtation by using greater laughter, asking more questions, and referring to the past. Even friendliness is signaled by greater laughter.

However, we do find gender differences. In some cases these seem to correspond to gender norms. When (perceived as) flirting, women raise their pitch, while men drop their pitch. Laughter is the most significant factor in women perceived as flirting; laughter for men is less highly weighted. Women make much more use of very quiet speech (possibly indicating private talk). But in other cases, there is no natural correspondence to gender norms. Flirting for men is accompanied by an expansion of pitch range.

Gender differences exist as well for the other variables. Women perceived as awkward are more likely to ask questions than the men. By a number of measures friendly men talk more than friendly women (the men have longer turns, more words, less backchanneling).

While the features for friendly and flirtatious speech overlap, there are clear differences. For men, flirtatious speech involves more questions, use of "you", lower pitch, and increased pitch range. Friendly speech involves higher pitch and decreased pitch range/variation. For women, flirtation speech seems to be more other-directed, involving more questions, NTRIs and the use of 'you', as well as higher pitch and quieter speech. Friendly speech involves no questions, more variable pitch, and is louder.

In addition to these implications for social science, our work has implications for the extraction of meaning in general. A key focus of our work was on ways to extract useful dialog act features (NTRIs, backchannels, appreciations, disfluencies) with very shallow methods. These features were indeed extractable and proved to be useful features in classification. Our results also introduce a new measure of pitch variation, the standard deviation across turns of pitch standard deviation within turns. This feature seemed to play as large a role as more traditional features like pitch range, suggesting the use of a broader arsenal of measure of pitch variation.

We are now extending these results to explore a wider variety of richer lexical features such as the sophisticated lexical features applied in LIWC (Pennebaker et al., 2007). In addition, we are attempting to predict date outcomes including 'clicking' or 'liking'. For this future task we will be able to control for other variables that are coded in our corpus such as age, perceived attractiveness, and demographic information and other aspects of personal background.

Acknowledgments

Removed for anonymity.

References

J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. 2002. Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog. In *INTERSPEECH-02*.

Paul Boersma and David Weenink. 2005. Praat: doing phonetics by computer (version 4.3.14). [Computer program]. Retrieved May 26, 2005, from `http://www.praat.org/`.

Susan E. Brennan and M. F. Schober. 2001. How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44:274–296.

Susan E. Brennan and M. Williams. 1995. The feeling of anothers knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34:383–398.

M. A. Cohn, M. R. Mehl, and J. W. Pennebaker. 2004. Linguistic markers of psychological change surrounding september 11, 2001. *Psychological Science*, 15:687–693.

Frank Enos, Elizabeth Shriberg, Martin Graciarena, Julia Hirschberg, and Andreas Stolcke. 2007. Detecting Deception Using Critical Segments. In *INTERSPEECH-07*.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL Labeling Project Coder's Manual, Draft 13. Technical Report 97-02, University of Colorado Institute of Cognitive Science.

Daniel Jurafsky. 2001. Pragmatics and computational linguistics. In Laurence R. Horn and Gregory Ward, editors, *Handbook of Pragmatics*. Blackwell.

C. M. Lee and Shrikanth S. Narayanan. 2002. Combining acoustic and language information for emotion recognition. In *ICSLP-02*, pages 873–876, Denver, CO.

Jackson Liscombe, Jennifer Venditti, and Julia Hirschberg. 2003. Classifying Subject Ratings of Emotional Speech Using Acoustic Features. In *INTERSPEECH-03*.

M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards. 2003. Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, 29:665–675.

Andrew Y. Ng. 2004. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *ICML 2004*.

J. W. Pennebaker and T. C. Lay. 2002. Language use and personality during crises: Analyses of mayor rudolph giulianis press conferences. *Journal of Research in Personality*, 36:271–282.

J. W. Pennebaker, R.E. Booth, and M.E. Francis. 2007. Linguistic inquiry and word count: Liwc2007 operators manual. Technical report, University of Texas.

Andrew Rosenberg and Julia Hirschberg. 2005. Acoustic/prosodic and lexical correlates of charismatic speech. In *EUROSPEECH-05*, pages 513–516, Lisbon, Portugal.

S. S. Rude, E. M. Gortner, and J. W. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18:1121–1133.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53:361–382.