High Frequency Word Entrainment in Spoken Dialogue

Ani Nenkova
Dept. of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
nenkova@seas.upenn.edu

Agustín Gravano
Dept. of Computer Science
Columbia University
New York, NY 10027, USA
agus@cs.columbia.edu

Julia Hirschberg
Dept. of Computer Science
Columbia University
New York, NY 10027, USA
julia@cs.columbia.edu

Abstract

Cognitive theories of dialogue hold that entrainment, the automatic alignment between dialogue partners at many levels of linguistic representation, is key to facilitating both production and comprehension in dialogue. In this paper we examine novel types of entrainment in two corpora—Switchboard and the Columbia Games corpus. We examine entrainment in use of *high-frequency words* (the most common words in the corpus), and its association with dialogue naturalness and flow, as well as with task success. Our results show that such entrainment is predictive of the perceived naturalness of dialogues and is significantly correlated with task success; in overall interaction flow, higher degrees of entrainment are associated with more overlaps and fewer interruptions.

1 Introduction

When people engage in conversation, they adapt the way they speak to their conversational partner. For example, they often adopt a certain way of describing something based upon the way their conversational partner describes it, negotiating a common description, particularly for items that may be unfamiliar to them (Brennan, 1996). They also alter their amplitude, if the person they are speaking with speaks louder than they do (Coulston et al., 2002), or reuse syntactic constructions employed earlier in the conversation (Reitter et al., 2006). This phenomenon is known in the literature as entrainment, accommodation, adaptation, or alignment.

There is a considerable body of literature which posits that entrainment may be crucial to human perception of dialogue success and overall quality, as well as to participants' evaluation of their conversational partners. Pickering and Garrod (2004) propose that the automatic alignment at many levels of linguistic representation (lexical, syntactic and semantic) is key for both production and comprehension in dialogue, and facilitates interaction. Goleman (2006) also claims that a key to successful communication is human ability to synchronize their communicative behavior with that of their conversational partner. For example, in laboratory studies of non-verbal entrainment (mimicry of mannerisms and facial expressions between subjects and a confederate), Chartrand and Bargh (1999) found not only that subjects displayed a strong unintentional entrainment, but also that greater entrainment/mimicry led subjects to feel that they liked the confederate more and that the overall interaction was progressing more smoothly. People who had a high inclination for empathy (understanding the point of view of the other) entrained to a greater extent than others. Reitter et al. (2007) also found that degree of entrainment in lexical and syntactic repetitions that occurred in only the first five minutes of each dialogue significantly predicted task success in studies of the HCRC Map Task Corpus.

In this paper we examine a novel dimension of entrainment between conversation partners: the use of *high-frequency words*, the most frequent words in the dialogue or corpus. In Section 2 we describe experiments on high-frequency word entrainment and perceived dialogue naturalness in Switchboard dia-

Proceedings of ACL-08: HLT, Short Papers (Companion Volume), pages 169–172,
Columbus, Ohio, USA, June 2008.

logues. The degree of high-frequency word entrainment predicts naturalness with an accuracy of 67% over a 50% baseline. In Section 3 we discuss experiments on the association of high-frequency word entrainment with task success and turn-taking. Results show that degree of high-frequency word entrainment is positively and significantly correlated with task success and proportion of overlaps in these dialogues, and negatively and significantly correlated with proportion of interruptions.

2 Predicting perceived naturalness

2.1 The Switchboard Corpus

The Switchboard Corpus (Godfrey et al., 1992) is a collection of recordings of spontaneous telephone conversations between speakers of many varieties of American English who were asked to discuss a preassigned topic from a set including favorite types of music or the new roles of women in society. The corpus consists of 2430 conversations with an average duration of 6 minutes, for a total of 240 hours and three million words. The corpus has been orthographically transcribed and annotated for degree of naturalness on Likert scales from 1 (very natural) to 5 (not natural at all).

2.2 Entrainment and perceived naturalness

Previous studies (Niederhoffer and Pennebaker, 2002) have suggested that adaptation in overall word count as well as words of particular parts of speech, or words associated with emotion or with various cognitive states, can predict the degree of coordination and engagement of conversational partners. Here, we examine conversational partners' similarity in high-frequency word usage in the Switchboard corpus as a predictor of the hand-annotated naturalness scores for their conversation. Using entrainment over the most frequent words in the entire corpus has the advantage of avoiding sparsity problems; we hypothesize that it will be more general and robust than attempting to measure lexical entrainment over the high-frequency words that occur in a particular conversation.

Our measure of entrainment $entr(w)$ is defined as the negated absolute value of the difference between the fraction of times a particular word w is used by the two speakers S_1 and S_2. More formally,

$$entr(w) = -\left|\frac{count_{S_1}(w)}{ALL_{S_1}} - \frac{count_{S_2}(w)}{ALL_{S_2}}\right|$$

Here, ALL_{S_i} is the number of all words uttered by speaker S_i in the given conversation, and $count_{S_i}(w)$ is the number of times S_i used word w.

The $entr(w)$ statistic was computed for the 100 most common words in the entire Switchboard corpus and feature selection was used to determine the 25 most predictive words used for later classification: *um, how, okay, go, I've, all, very, as, or, up, a, no, more, something, from, this, what, too, got, can, he, in, things, you, and*.

The data for the experiments was a balanced set of 250 conversations rated "1" (very natural) and 250 examples of problematic conversations with ratings of 3, 4 or 5. The accuracy of predicting the binary naturalness (ratings of 1 or 3-5) of each conversation from a logistic regression model is 63.76%, significantly over a 50% random baseline. This result confirms the hypothesis that entrainment in high-frequency word usage is a good indicator of the perceived naturalness of a conversation.

Some of our 25 high-frequency words are in fact *cue phrases*, which are important indicators of dialogue structure. This suggests that a more focused examination of this class of words might be useful.

3 Association with task success and dialogue flow

3.1 The Columbia Games Corpus

The Columbia Games Corpus (Benus et al., 2007) is a collection of 12 spontaneous task-oriented dyadic conversations elicited from native speakers of Standard American English. Subjects played a series of computer games requiring verbal communication between partners to achieve a common goal, either identifying matching cards appearing on each of their screens, or moving an object on one screen to the same location in which it appeared on the other, where each subject could see only their own screen. The games were designed to encourage frequent and natural conversation by engaging the subjects in competitive yet collaborative tasks. For example, players could receive points in the games in a variety of ways and had to negotiate the best strategy

for matching cards; in other games, they received more points if they could place objects in exactly the same location. Subjects were scored on each game and their overall score determined the additional monetary compensation they would receive. A total of 9h 8m ($\sim$73,800 words) of dialogue were recorded. All files in the corpus were orthographically transcribed and words were hand-aligned by trained annotators. A subset of the corpus was also labeled for different types of turn-taking behavior. These include *(i)* **smooth turn exchanges**—speaker S_2 takes the floor after speaker S_1 has completed her turn, with no overlap; *(ii)* **overlaps**—S_2 starts his turn before S_1 has completely finished her turn, but S_1 does complete her turn; *(iii)* **interruptions**—S_2 starts talking before S_1 completes her turn, and as a result S_1 does not complete her utterance. We used these annotations to study the association between entrainment and turn-taking behavior.

3.2 Entrainment and task success

In the Columbia Games Corpus, we hypothesize that the game score achieved by the participants is a good measure of the effectiveness of the dialogue. To determine the extent to which task success is related to the degree of entrainment in high-frequency word usage, we examined 48 dialogues. We computed the correlation coefficient between the game score (normalized by the highest achieved score for the game type) and two different ways of quantifying the degree of entrainment between the speakers (S_1 and S_2) in several word classes. In addition to overall high-frequency words, we looked at two subclasses of words often used in dialogue:

25MF-G The 25 most frequent words in the game.

25MF-C The 25 most frequent words over the entire corpus: *the, a, okay, and, of, I, on, right, is, it, that, have, yeah, like, in, left, it's, uh, so, top, um, bottom, with, you, to*.

ACW Affirmative cue words: *alright, gotcha, huh, mm-hm, okay, right, uh-huh, yeah, yep, yes, yup*. There are 5831 instances in the corpus (7.9% of all words).

FP Filled pauses: *uh, um, mm*. The corpus contains 1845 instances of filled pauses (2.5% of all tokens).

We generalize our measure of word entrainment $entr(w)$ to each of these *classes* of words c:

$$ENTR_1(c) = \sum_{w \in c} entr(w)$$

$ENTR_1$ ranges from 0 to $-\infty$, with 0 meaning perfect match on usage of lexical items in class c. An alternative measure of entrainment that we experimented with is defined as

$$ENTR_2(c) = -\frac{\sum_{w \in c} |count_{S_1}(w) - count_{S_2}(w)|}{\sum_{w \in c} (count_{S_1}(w) + count_{S_2}(w))}$$

The entrainment score defined in this way ranges from 0 to -1, with 0 meaning perfect match on lexical usage and -1 meaning perfect **mis**match.

The correlations between the normalized game score and these measures of entrainment are shown in Table 1. $ENTR_1$ for the 25 most frequent words, both corpus-wide and game-specific, is highly and significantly correlated with task success, with stronger results for game-specific words. For the

	$ENTR_1$		$ENTR_2$	
Word class	*cor*	*p*	*cor*	*p*
25MF-C	**0.341**	**0.018**	0.187	0.202
25MF-G	**0.376**	**0.008**	0.260	0.074
ACW	0.230	0.116	**0.372**	**0.009**
FP	-0.080	0.591	-0.007	0.964

Table 1: Pearson's correlation with game score.

filled pauses class, there is essentially no correlation between entrainment and task success, while for affirmative cue words there is association only under the $ENTR_2$ definition of entrainment. The difference in results between $ENTR_1$ and $ENTR_2$ suggests that the two measures of entrainment capture different aspects of dialogue coordination and that exploring various formulations of entrainment deserves future attention.

3.3 Dialogue coordination

The coordination of turn-taking in dialogue is especially important for successful interaction. Speech overlaps (O), might indicate a lively, highly coordinated conversation, with participants anticipating the end of their interlocutor's speaking turn. Smooth switches of turns (S) with no overlapping speech are also characteristic of good coordination, in cases where these are not accompanied by long pauses between turns. On the other hand, interruptions (I) and long inter-turn **latency** (L)—long simultaneous pauses by the speakers— are generally perceived as a sign of poorly coordinated dialogues.

To determine the relationship between entrainment and dialogue coordination, we examined the correlation between entrainment types and the proportion of interruptions, smooth switches and overlaps, for which we have manual annotations for a subset of 12 dialogues. We also looked at the correlation of entrainment with mean latency in each dialogue. Table 2 summarizes our major findings.

		cor	*p*
$ENTR_1$(25MF-C)	I	$-$**0.612**	**0.035**
$ENTR_1$(25MF-G)	I	$-$0.514	0.087
$ENTR_1$(ACW)	O	**0.636**	**0.026**
$ENTR_2$(ACW)	O	**0.606**	**0.037**
$ENTR_1$(FP)	O	**0.750**	**0.005**
$ENTR_2$(25MF-G)	O	**0.605**	**0.037**
$ENTR_2$(25MF-G)	S	$-$**0.663**	**0.019**
$ENTR_2$(ACW)	L	$-$**0.757**	**0.004**
$ENTR_2$(25MF-G)	L	$-$0.523	0.081

Table 2: Pearson's correlation with proportion of overlaps, interruptions, smooth switches, and mean latency.

The two measures that were significantly correlated with task success—$ENTR_1$(25MF-C) and $ENTR_1$(25MF-G)—also correlated *negatively* with the proportion of interruptions in the dialogue. This finding could have important implications for the development of spoken dialog systems (SDS). For example, a measure of entrainment might be used to anticipate the user's propensity to interrupt the system, signalling the need to change dialogue strategy. It also suggests that if the system entrains *to users* it might help to reduce such interruptions. While our study is of association, not causality, this suggests future areas of investigation.

Our other correlations reveal that turn exchanges characterized by overlaps are reliably associated with entrainment in usage of affirmative cue word, filled pauses and game-specific most frequent words. Long latency is negatively associated with entrainment in affirmative cue words and game-specific most frequent words. Overall, the more entrainment, the more engaged the participants and the better coordination there is between them, with shorter latencies and more overlaps.

Unexpectedly, smooth switches correlate negatively with entrainment in game-specific most frequent words. This result might be confounded by the presence of long latencies in some switches. While smooth switches are desirable, especially in SDS, long latencies between turns can indicate lack of coordination.

4 Conclusion

We present a corpus study relating dialogue naturalness, success and coordination with speaker entrainment on common words: most frequent words overall, most frequent words in a dialogue, filled pauses, and affirmative cue words. We find that degree of entrainment with respect to most frequent words can distinguish dialogues rated most natural from those rated less natural. Entrainment over classes of common words also strongly correlates with task success and highly engaged and coordinated turn-taking behavior. Entrainment over corpus-wide most frequent words significantly correlates with task success and minimal interruptions—important goals of SDS. In future work we will explore the consequences of system entrainment to SDS users in helping systems achieve these goals, and the use of simple measures of entrainment to modify dialogue strategies in order to decrease the occurrence of user interruptions.

Acknowledgments

This work was funded in part by NSF IIS-0307905.

References

S. Benus, A. Gravano, and J. Hirschberg. 2007. The prosody of backchannels in American English. *ICPhS'07*.

S.E. Brennan. 1996. Lexical entrainment in spontaneous dialog. *ISSD'96*.

T. Chartrand and J. Bargh. 1999. The chameleon effect: the perception-behavior link and social interaction. *J. of Personality & Social Psych.*, 76(6):893–910.

R. Coulston, S. Oviatt, and C. Darves. 2002. Amplitude convergence in children's conversational speech with animated personas. *ICSLP'02*.

J. Godfrey, E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *ICASSP'92*.

Daniel Goleman. 2006. *Social Intelligence*. Bantam.

K. Niederhoffer and J. Pennebaker. 2002. Linguistic style matching in social interaction.

M. J. Pickering and S. Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

D. Reitter and J. Moore. 2007. Predicting success in dialogue. *ACL'07*.

D. Reitter, F. Keller, and J.D. Moore. 2006. Computational Modelling of Structural Priming in Dialogue. *HLT-NAACL'06*.