Temporal Distributional Analysis

Nigel G. Ward
Computer Science Department
University of Texas at El Paso
El Paso, Texas, 79968 USA
nigelward@acm.org

Abstract

Two salient characteristics of spoken dialogs, in contrast to written texts, is that they are processes in time and that they are co-constructed by the interlocutors. Most current corpus-based methods for analyzing dialog phenomena, however, abstract away from these characteristics. This paper introduces a new corpus-based analysis method, temporal distributional analysis, which can reveal such aspects of dialog. Given a word of interest, this method identifies which words tend to co-occur with it at specific temporal offsets. This can be done not only for words produced by the same speaker but also for the interlocutor's words. This paper explains the method, presents several ways to visualize the results, illustrates what it reveals about the words *I*, *uh* and *uh-huh*, compares it to non-temporal distributional analysis, and discusses potential applications to speech recognition, generation, and synthesis.

1 Introduction

Although spoken dialog is fundamentally different from written language in several ways, it is common for dialog researchers to work with textual representations. Although convenient, this can lose useful information. This paper addresses this problem with a new type of distributional analysis; a new member of the widely used family of techniques implementing Firth's well-known maxim that "a word is known by the company it keeps." In particular, this paper looks at words as events in time: rather than merely examining what neighbors a word has, it considers *when* those neighbors occurred, that is, their timing relative to the word of interest. It also examines how words by a speaker relate temporally to the words of the interlocutor.

In general, in studies of language use, the unit of analysis has been the word, although psycholinguists more commonly use the elapsed second, as a critical variable in studies of reactions, perceptions, and responses. For dialog, although time is of the essence (Clark, 2002), most researchers still tend to still work in terms of sequences of words, although there are notable exceptions, including (Bard et al., 2002; Boltz, 2005; Ji and Bilmes, 2004), and, non-quantitatively, many practitioners of Conversation Analysis methods. Existing methods for studying dialog dynamics are, however, far from suitable for general use, all having one or more weaknesses, including being impressionistic, of limited use, theory-bound, or labor-intensive. Thus there is a need for general methods for studying the temporal aspects of dialog; and this paper presents one.

Section 2 introduces the method and its implementation; Section 3 illustrates the application of the method to some common words and some ways to visualize the results; Section 4 considers the value of the method; and Section 5 discusses possible applications and future work.

2 Definitions

Figure 1 illustrates how words might occur in time in a dialog. Clearly here *very* is part of the context of *cat*, but there are various possible ways to more specificaly characterize the relative position of the two. For example, one might simply tag *very* as oc-

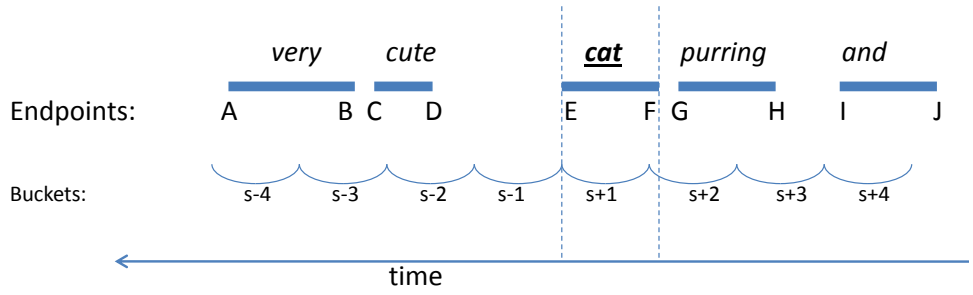

Figure 1: Illustration of words occurring in time, with *cat* taken as the word of interest and the others the context.

curring as the second word before *cat*.

The new idea here is to pay attention to the temporal relation between the two. While this would bring no new information if every word had the same duration and there were no pauses between words, in fact real spoken dialog does have pauses of various lengths and speaking rate variations, and these variations are often indicative of cognitive state and information state, and relate to the words that have appeared and that are likely to appear (Goldman-Eisler, 1967; Bell et al., 2009; Ward et al., 2011).

The temporal relation between the occurrences of two words can be measured in various ways. One metric would be the time between onsets, A-E in the example. However since a word, once initiated, can be stretched out at will to dovetail with the next word, or to establish a "rhythm," or to otherwise help the listener predict the upcoming words, it seems probably more useful to use instead the distance from the end, here B-E. For words after the word of interest, for example *and*, the metric is similarly the difference from the end time of the word of interest to the onset of the context word: F-I in the example.

These metrics also work for words in the interlocutor's track.

To identify the words which occur frequently in various temporal relations to the word of interest, we can count, over the entire corpus, for all occurrences of *cat*, say, which words are more frequent at certain time offsets. For convenience these are discretized, as suggested in the figure by the buckets. Thus, for example, the B-E distance for *very* falls in bucket *s-3*. For distances relative to the end of the word of interest a similar set of buckets, not shown, is used.

From the counts over the whole corpus, we can compute the degree to which a context word x is characteristic of a certain bucket for a word of interest. In particular, this can be done by comparing the in-bucket probability to the overall (unigram) probability for x. For example, we can compute the ratio of the probability of *very* appearing in bucket *s-3* to the probability of *very* appearing anywhere in the corpus. This we call the R ratio (Ward et al., 2011). From the probability of each word in each bucket, the "bucket probability," that is, its count in the bucket for t divided by the total in that bucket,

$$P_{tb}(w_i@t) = \frac{count(w_i@t)}{\sum_j count(w_j@t)} \quad (1)$$

we can compute the ratio of this to the standard unigram probability:

$$R(w_i@t) = \frac{P_{tb}(w_i@t)}{P_{unigram}(w_i)} \quad (2)$$

If R is 1.0 there is no connection and no mutual information; larger values of R indicate positive correlations, and lower values of R indicate words that are rare in a given context position.

Although this paper looks only at individual words in the context, independently of each other, the method could also be applied to contextual word pairs or ngrams.

In the tables and figures below, these R-ratios were computed over a 650K word subset of Switchboard, a corpus of unstructured two-party telephone conversations among strangers (ISIP, 2003). To test whether a R-ratio is significantly different from 1.0, the chi-squared test can be applied, where the null hypothesis is that the context word occurs in a certain bucket as often as expected from the unigram probability of the word and the total number of

	Preceding Buckets						Following Buckets					
	8-6	6-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-6	6-8
I:	$1.68^{\#}$	$1.74^{\#}$	$1.96^{\#}$	$2.20^{\#}$	$2.26^{\#}$	$2.05^{\#}$	$1.36^{\#}$	$1.61^{\#}$	$1.75^{\#}$	$1.66^{\#}$	$1.50^{\#}$	$1.47^{\#}$
you:	...	$...^{\#}$	...	$...^{\#}$	$0.76^{\#}$	...	$0.59^{\#}$	$0.75^{\#}$	$...^{\#}$	...	$...^{*}$	$...^{\#}$
it:	$...^{\#}$	$1.22^{\#}$	$1.27^{\#}$	$1.27^{\#}$	$1.22^{\#}$	$...^{*}$	$...^{\#}$	$1.59^{\#}$	$1.29^{\#}$	$1.24^{\#}$	$1.29^{\#}$	$1.35^{\#}$
that:	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	$1.53^{\#}$	...	$1.64^{\#}$	$1.33^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$
the:	$...^{\#}$	$...^{\#}$	$...^{\#}$	...	$...^{\#}$	$0.56^{\#}$	$0.70^{\#}$	$1.21^{\#}$	$1.21^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$
a:	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	...	$0.29^{\#}$	$...^{\#}$	$1.57^{\#}$	$1.39^{\#}$	$1.27^{\#}$	$1.22^{\#}$	$...^{\#}$
and:	$1.22^{\#}$	$1.23^{\#}$	$1.22^{\#}$	$1.23^{\#}$	$1.28^{\#}$	$2.32^{\#}$	$0.15^{\#}$	$0.59^{\#}$	...	$1.24^{\#}$	$1.32^{\#}$	$1.38^{\#}$
but:	$1.23^{\#}$	$1.29^{\#}$	$1.45^{\#}$	$1.71^{\#}$	$1.95^{\#}$	$3.51^{\#}$	$0.15^{\#}$	...	$1.62^{\#}$	$1.54^{\#}$	$1.48^{\#}$	$1.43^{\#}$
to:	$1.20^{\#}$	$1.20^{\#}$	$...^{\#}$	$...^{\#}$	...	$0.45^{\#}$	$1.28^{\#}$	$1.47^{\#}$	$1.37^{\#}$	$1.29^{\#}$	$1.26^{\#}$	$1.23^{\#}$
of:	$...^{\#}$	$...^{\#}$	$...^{\#}$	...	$...^{\#}$	$0.62^{\#}$	$0.48^{\#}$	$1.41^{\#}$	$1.26^{\#}$	$1.21^{\#}$	$1.23^{\#}$	$...^{\#}$
yeah:	$...^{+}$	$...^{\#}$	$...^{\#}$	$1.21^{\#}$	$1.34^{\#}$	$2.33^{\#}$	$0.07^{\#}$	$0.12^{\#}$	$0.18^{\#}$	$0.31^{\#}$	$0.48^{\#}$	$0.62^{\#}$
so:	$...^{\#}$	$...^{\#}$	$...^{\#}$	$1.24^{\#}$	$1.44^{\#}$	$2.79^{\#}$	$0.47^{\#}$	$0.52^{\#}$	...	$...^{\#}$	$...^{\#}$	$...^{\#}$
laughter:	$...^{\#}$	$...^{+}$	...	...	...	$...^{\#}$	$0.28^{\#}$	$0.64^{\#}$	$0.82^{\#}$	$0.81^{\#}$	$...^{\#}$	$...^{\#}$
well:	$1.25^{\#}$	$1.23^{\#}$	$1.36^{\#}$	$1.67^{\#}$	$1.92^{\#}$	$4.08^{\#}$	$0.31^{\#}$	$0.31^{\#}$	$0.44^{\#}$	$0.50^{\#}$	$0.57^{\#}$	$0.71^{\#}$
uh:	$...^{\#}$	$...^{\#}$	$...^{\#}$	$1.27^{\#}$	$1.38^{\#}$	$1.53^{\#}$	$0.50^{\#}$	$0.81^{\#}$	...	$...^{\#}$	$1.20^{\#}$	$...^{\#}$
uh-huh:	$0.56^{\#}$	$0.56^{\#}$	$0.49^{\#}$	$0.35^{\#}$	$0.26^{\#}$	$0.18^{\#}$	$0.01^{\#}$	$0.02^{\#}$	$0.04^{\#}$	$0.13^{\#}$	$0.28^{\#}$	$0.36^{\#}$
know:	$1.25^{\#}$	$1.30^{\#}$	$1.33^{\#}$	$1.32^{\#}$	$1.23^{\#}$	$2.22^{\#}$	$2.80^{\#}$	...	$...^{\#}$	$1.30^{\#}$	$1.32^{\#}$	$1.37^{\#}$
think:	$1.25^{\#}$	$1.27^{\#}$	$1.27^{\#}$	$1.40^{\#}$	$1.43^{\#}$	$1.43^{\#}$	$10.56^{\#}$	$1.36^{\#}$	$1.26^{\#}$	$...^{*}$	$...^{*}$	...
OOS:	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	$0.82^{\#}$	...	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$	$...^{\#}$

Figure 2: R-ratios for common words in the vicinity of *I*. The "Preceding Bucket 8-6" column, for example, is for occurrences of words ending more than 6 but less than 8 seconds before the start of an occurrence of *I*, and similarly for the others. Only values interestingly different from 1.00 are shown: those whose r-ratio is greater than 1.2 or less than 0.83. The trailing symbols indicate significance: $+$ indicates $p < .05$, * $p < .02$, and # $p < .01$.

words in that bucket, where the sample population is relative to all occurrences of the word of interest in the corpus.

3 Illustrations and Observations

This section presents some raw data, using several visualization methods, and some observations about the distributions and possible underlying causes.

Table 2 shows R-ratios for some words as they appear in various buckets relative to 26293 occurrences of the word *I*. The context words shown were chosen as the 10 most frequent words, some common discourse markers, and the words *think* and *know*. Values for the class of all other words are shown as *OOS* (out of shortlist). The assymmetry for *I* as a context word is due to the differing total counts in the various buckets.

The alternative representation seen in Figure 3 uses the vertical positioning of words to indicate their R-ratios. For conciseness, in each bucket the only words shown are those whose ratios are above 1.4 or below 0.7, and where the difference from 1 is significant at $p < .01$. Within each cell, words are ordered to show syntactic and semantic similarities.

Another alternative representation, Figure 4, highlights how the R-ratios vary over time.

In this data some interesting patterns are seen. For example, the word *but* is more common than usual starting around 1 second after the word *I*; in contrast *and* doesn't become more frequent until around 4 seconds later. This difference may reflect the tendency inl conversation to not let a partially true statement about oneself stand for more than a couple of seconds before giving the caveat.

	preceding					following				
R	8-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-8
> 4.0					well	think				
> 2.8					but	know				
> 2.0			I	I	I, and, so, know, yeah					
> 1.4	I	I, but	but, well	but, well, so, think	think, uh, that	I, it, that, a to	I, but	I, but	I, but	I, but
$< .71$	uh-huh				the	the, you	and		well	yeah, uh
$< .50$		uh-huh	uh-huh		of	of, so		well		
$< .35$				uh-huh	a, to	well, um, laughter	well		yeah	uh-huh
$< .25$					uh-huh	yeah, and, but, uh-huh,	uh-huh, yeah	uh-huh, yeah	uh-huh	

Figure 3: Same-speaker words with notably high and low R-ratios around the word *I* as the word of interest.

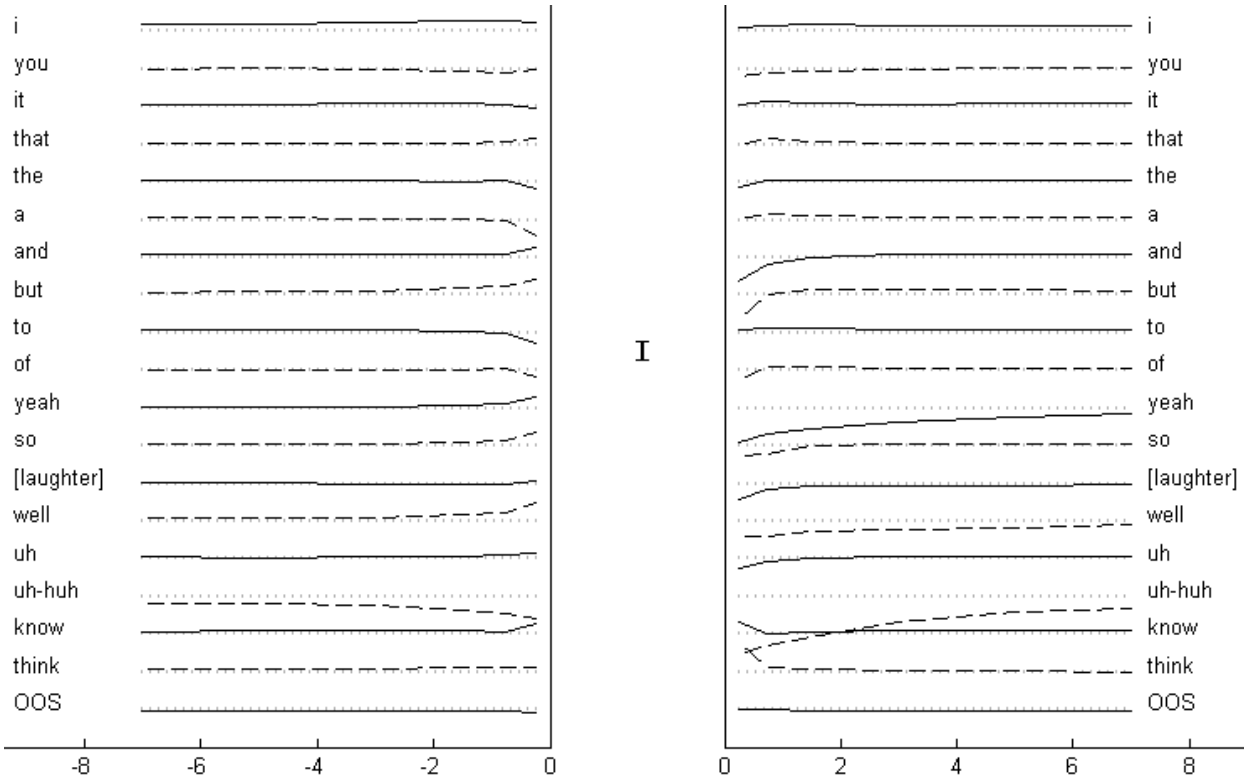

Figure 4: Log R-ratios as a function of time for various words in the vicinity of *I*. The dotted lines indicate the baseline (R=1).

R	8-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-8
> 4.0				uh-huh, laughter	uh-huh, laughter, yeah	yeah, uh-huh, laughter	uh-huh, yeah, laughter	uh-huh, yeah, laughter	uh-huh, yeah	
> 2.8			uh-huh, laughter	yeah			well	well	well, laughter	uh-huh, yeah
> 2.0	laughter	yeah, laughter, uh-huh	yeah							laughter
> 1.4	yeah, uh-huh	you		so	so	I, you, it, think, know, well	I, think	I, think	I, think	I, think
$< .71$			and, uh	the, a, and, uh	and, of	to	to, the	a, and, to, of	and	
$< .50$					the, a, to		of			

Figure 5: Interlocutor words with notably high and low R-ratios in the vicinity of *I*.

Figure 5 shows the results when the context words are taken from the interlocutor's track. There is a strong tendency for *I* to co-occur near *uh-huh*, *[laughter]* and *yeah* by the interlocutor, and also a tendency for occurrences of *I* to be followed by the word *I* by the interlocutor.

The contexts of the word *uh* are seen in Figures 6 and 7. These show that *uh* frequently closely follows an *and* or *but* by the speaker, and *uh-huh*, *yeah*, and *[laughter]* by the interlocutor; and that it is frequently closely followed by *I*, *know*, and *you* by the speaker, and by *yeah, well, [laughter]*, and *but* by the interlocutor, presumably reflecting feedback and turn-grab actions.

The context words spoken by the interlocutor in the vicinity of *uh-huh* are seen in Figure 8. Among the interesting patterns seen is the relation with *I*: *uh-huh* is often preceded by a word *I* by the interlocutor 4–8 seconds earlier, counter-indicated by an *I* less than one second earlier, but commonly followed with an *I* within 1 second. Perhaps this reflects a dialog pattern where an initial *I* is typically followed by some new information, then by feedback from the listener, then very swiftly by another *I* introducing more information; although there are probably also deeper explanations involving syntactic, semantic, pragmatic, and cognitive chunking and response time factors.

4 The Value of Temporal Distributional Analysis

The identification of previously unknown regularities in dialog, above, suggests that this method is valuable. However, as a proposed advance, it is necessary to consider whether it really is an improvement over non-temporal methods.

The most direct comparison is to look at which words co-occur with the word of interest across spans measured, not in seconds, but in words. Figure 9 is an example, showing the pattern of contextual co-occurring words by the same speaker in the vicinity of occurrences of *I*, limited to the most frequent 10 words for conciseness. In generating this figure pauses were ignored, even long ones that might typically be thought to reset the context; this allowed long-distance patterns to appear, in particular for words commonly preceded or followed by silence, such as *uh-huh*. (While on the topic of silence, I note that the method presumes that silence is nothing more than a device to let some time go by; but in some cases it may have more specific meanings, and one might try treating silences of various durations differently, perhaps as functioning as different context "words.")

Comparing this with Figure 3, all the common patterns there are also seen here, and this was true also for the 17 other common words and discourse markers I looked at. Thus, the hope of finding new

R	8-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-8
> 4.0					but, and					
> 2.8										
> 2.0										
> 1.4	uh	uh	uh	uh	that, so well, think	I		uh		
$< .71$	laughter	laughter	laughter	laughter	you, a, yeah, laughter	that	but, and			laughter, well
$< .50$	uh-huh				laughter	and, to yeah, laughter	well	well, laugh-ter	laughter, well	yeah
$< .35$										yeah
$< .25$		uh-huh	uh-huh	uh-huh	uh-huh	uh-huh	yeah, uh-huh, laughter	uh-huh, yeah	uh-huh	uh-huh

Figure 6: Same-speaker words with notably high and low R-ratios in the vicinity of *uh*.

R	8-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-8
> 4.0		uh-huh	uh-huh	uh-huh, yeah, laughter	uh-huh, yeah, laughter		uh-huh	uh-huh, yeah	uh-huh, yeah	uh-huh
> 2.8	uh-huh		yeah, laughter			uh-huh, yeah	yeah	well	well	yeah
> 2.0	laughter	laughter, yeah				well, laughter	well, laughter	laughter	laughter	well, laughter
> 1.4	yeah	you, so	you	you	well	you, but	that		I, think	I, think
$< .71$		and	and, uh	uh, I, the, a, of, know	and, know	uh, and	and, a	and, to, but	and, but	
$< .50$				and, to						
$< .35$					the, to					

Figure 7: Interlocutor words with notably high and low R-ratios in the vicinity of *uh*.

R	8-4	4-2	2-1	1-.5	.5-0	0-.5	.5-1	1-2	2-4	4-8
> 2.8						and, so				
> 2.0						but				
> 1.4	I		a, to	a, the, of	it	uh, I	and, but, I	and	and, of	and, so
$< .71$		yeah		I, you, so, uh, know		of	well	well	well	well, uh-huh
$< .50$	uh-huh	laughter		but, think		yeah, well, laughter	laughter	laughter		yeah
$< .35$			laughter, yeah, well	laughter, well				yeah	uh-huh, yeah	
$< .25$		uh-huh	uh-huh	uh-huh, yeah	uh-huh	uh-huh	uh-huh, yeah	uh-huh		

Figure 8: Interlocutor words with notably high and low R-ratios in the vicinity of *uh-huh*.

R	–5	–4	–3	–2	–1	+1	+2	+3	+4	+5
> 2.8					and					
> 2.0	I		I	I	that, uh		I, it, to			
> 1.4		I, well		uh			that, a	I, that, a, it, to	I, a, that of	I, a, that
$< .71$				to						
$< .50$				a		uh		and		
$< .35$							and, of			
$< .25$					the, a, of, to	to, of, the, a, it, you, that				

Figure 9: Words with notably high and low R-ratios at various offsets from *I*. The –5 column indicates words occurring 5 words before *I*, and so on.

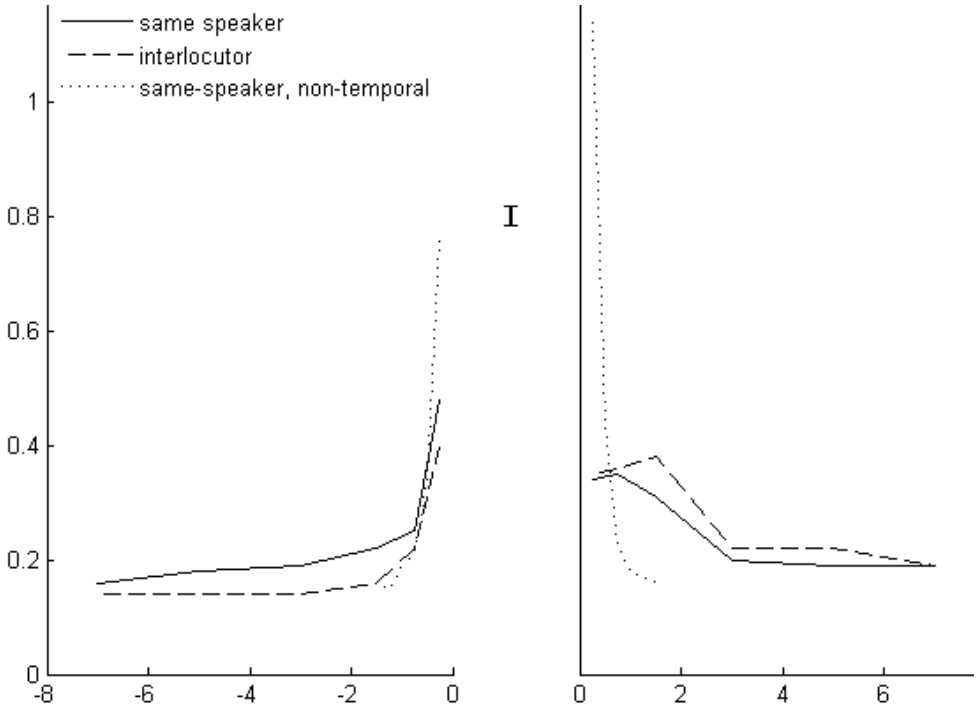

Figure 10: Per-bucket information as a function of time.

patternings by using time as the dependent variable was not fulfilled (looking only at same-speaker patternings, and only over relatively short distances).

Another way to compare is to estimate quantitatively the amount of information provided. The figures above show a general tendency for the R-ratios to become less extreme as the distance from the word of interest increases; this is to be expected, as a word likely to relate more to its closer neighbors. This suggests that temporal models may have greater value for longer distances, compared to standard sequential models, which might do well only for syntactic and similar effects which are strong over short distances.

Evaluating this proposition requires a way to estimate the informativeness of the various models. Building on the observation that more extreme R-ratios are more informative, and borrowing from Information Theory the use of the logarithm of the probability as measure of information content, the total information content in each bucket y of a model of the context of word w can be estimated as

$$A_{wy} = \sum_x P_x |logR| \quad (3)$$

where the informativeness of each R-ratio is weighted by the overall frequency of the associated context word x in the corpus. To properly apply this metric, one would need to vary not only the context word but also the word of interest across all the words in the corpus, and when doing so properly deal with sparseness.

As an illustrative example, Figure 10 shows the informativeness per bucket only for the word *I*, and computed only over the 18 context words seen above. The figure thus shows the A_{Iy} as a function of time for both the same speaker's context words (solid line) and the interlocutor's context words (dashed line), and in addition for the same speaker's context words as a function of distance in words (dotted line). The x-axis is in seconds: for the temporal buckets the informativeness is plotted at the bucket center; and for the distance-in-words buckets at the approximate average corresponding temporal offset, assuming for convenience that words average a quarter-second in length (Yuan et al., 2006) and ignoring the effect of pauses.

The figure suggests that measuring distance in seconds, not words, has more value for the more distant context, at least for the word *I*. The figure

also suggests that the word *I* relates somewhat more tightly to the words of the same speaker in the previous context, but more to the words of the interlocutor in the following context.

5 Discussion

This exploration has shown that indeed there are interesting temporal distributional patterns, both relative to the words by the same speaker, and relative to words by the interlocutor. This section discusses possible uses for this knowledge.

One is speech recognition, where good language models are essential. Identifying which words are likely to occur at certain positions in dialog should be able to help this, but I do not know whether these patterns are non-redundant to those provided by ngrams, dialog-act-based modeling or conditioning on times relative to turn-taking events (Shriberg et al., 1998; Ward et al., 2011).

Another reason to be interested in such patterns is for what they say about words. Detailed case studies of the properties of individual words are often a first step to linguistic insight, but common corpus-based methods generally reveal only syntactic and semantic properties. As a way to get at more elusive dialog and pragmatic properties, temporal distributional analysis may be widely useful; to this end I hope to create a web resource to support perusal of the temporal distributional patterns for any word of interest. Apart from scientific curiosity, these patterns may be useful for finding new dimensions of lexical similarity, where two words are similar if their configurations of frequent neighbors are similar. New aspects of similarity may support better methods for dimensionality reduction for the lexicon, which in turn is critical for tasks from language modeling to information retrieval.

Regardless of the existence or non-existence of deep, satisfying explanations for these patterns, they are real. This suggests that generated and synthesized speech for use in dialog should respect these patterns to be perceived as natural, and so such patterns may provide an additional, useful, constraint on the timings of words in dialog, especially in cases where cross-speaker effects, such as in turn-taking and "sub-utterance" phenomena, are important (Buss and Schlangen, 2010).

Acknowledgments

This work was supported in part by NSF Award IIS-0914868. I thank Justin McManus for discussion and the anonymous reviewers for comments.

References

Bard, E. G., Aylett, M. P., and Lickley, R. J. (2002). Towards a psycholinguistics of dialogue: Defining reaction time and error rate in a dialogue corpus. In Bos, J., Foster, M., and Matheson, J., editors, *EDILOG 2002: 6th workshop on the semantics and pragmatics of dialogue*, pages 29–36.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60:92–111.

Boltz, M. (2005). Temporal dimensions of conversational interaction: The role of response latencies and pauses in social impression formation. *Journal of Language and Social Psychology*, 24:103–138.

Buss, O. and Schlangen, D. (2010). Modelling sub-utterance phenomena in spoken dialogue systems. In *Proceedings of SemDial 2010 (Pozdial)*.

Clark, H. H. (2002). Speaking in time. *Speech Communication*, 36:5–13.

Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech*, 10:122–132.

ISIP (2003). Manually corrected Switchboard word alignments. Mississippi State University. Retrieved 2007 from http://www.ece.msstate.edu/research/isip/projects/switchboard/.

Ji, G. and Bilmes, J. (2004). Multi-speaker language modeling. In *Conference on Human Language Technologies*.

Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., and Van Ess-Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, 41:439–487.

Ward, N. G., Vega, A., and Baumann, T. (2011). Prosodic and temporal features for language modeling for dialog. *Speech Communication*. to appear.

Yuan, J., Liberman, M., and Cieri, C. (2006). Towards and integrated understanding of speaking rate in conversation. In *ICSLP*.