Measuring Adaptation Between Dialogs

Svetlana Stenchikova and **Amanda Stent**
Computer Science Department
Stony Brook University
Stony Brook, NY 11794-4400
sveta, stent@cs.sunysb.edu

Abstract

The paper proposes two new approaches for measuring adaptation between dialogs. These approaches permit measurement of adaptation both to conversational partner (*partner adaptation*) and to the local dialog context (*recency adaptation*), and can be used with different types of feature. We used these measures to study adaptation in the Maptask corpus of spoken dialogs. We show that for syntactic features, recency adaptation is stronger than partner adaptation; however, we find no significant differences for lexical adaptation using these measures.

1 Introduction

Numerous psycholinguistic studies have demonstrated that people adapt their language use in conversation to that of their conversational partners. For example, conversational partners adapt to each other's choice of words, particularly referring expressions (Brennan and Clark, 1996), converge on certain syntactic choices (Pickering et al., 2000; C. Lockridge, 2002), adapt their prosody to help their partners disambiguate syntactic ambiguities (Kraljic and Brennan, 2005), and also adapt using audiovisual information (Kraut et al., 2003).

Some of these results have been duplicated using corpus studies; for example, researchers have found evidence of within-speaker and between-speaker convergence to certain syntactic constructions (Dubey et al., 2006; Reitter et al., 2006). Corpus studies can be a good addition to more tightly controlled empirical studies in cases where there is a corpus already available. Corpus studies can confirm the results of psycholinguistic research, and can identify issues that may 'muddy' empirical results.

Finally, there is some evidence that people adapt their language use in conversation with computer partners. For example, researchers have shown that users of dialog systems adapt the system's choice of referring expressions (Brennan, 1996), the system's choice of modality for referring (Bell et al., 2000; Skantze, 2002), or the system's choice of words (Gustafson et al., 1997).

Currently, there is a debate in the psycholinguistics community about whether this adaptation is:

- *partner adaptation* – adaptation based on a model of the partner. This type of adaptation is sometimes called entrainment or audience design (Brennan and Clark, 1996; Horton and Gerrig, 2002).

- *recency adaptation* – adaptation due to the representations of words, concepts etc. being *activated*, or brought to the forefront during language production, by previous perception or comprehension. This type of adaptation is sometimes called convergence, priming or alignment (Brown and Dell, 1987; Pickering and Garrod, 2004; Chartrand and Bargh, 1999).

In this paper, we consider measures used in corpus-based studies of adaptation such as (Dubey et al., 2006; Reitter et al., 2006; Church, 2000). These measures do not permit examination of whether adaptation is due to the partner or to recency, and do

not measure the strength of adaptation. We propose two new measures, one that measures the presence of adaptation and another that measures its strength. Together, these measures can identify adaptation within a single document or between documents; can identify the strength of adaptation as well as its presence; and can be used to identify the source of the adaptation. We use these measures to study adaptation in the Maptask spoken dialog corpus. We show that for syntactic features, recency adaptation is stronger than partner adaptation; however, we find no significant differences for lexical adaptation using these measures. We close with some ideas about how to apply these measures to dialog system development, and some ideas for future work.

2 Other Measures

Church (Church, 2000) introduced a method for measuring lexical "adaptation" in text. This method determines whether appearance of a lexical feature in the 'priming portion' of a document affects the likelihood of its appearance in the 'target' (later) portion. This method requires the construction of a contingency table for each feature in a corpus of texts, showing how many of the texts contained the feature: (a) in the 'priming portion' only, (b) in the 'target' only, (c) in both portions, and (d) in neither portion. The probability of positive adaptation is computed as $c/(a+c)$. This must be compared with a prior probability, which is $(a+c)/(a+b+c+d)$. Church applied this method to the study of a corpus of text documents, treating the first half of each document as the 'priming portion' and the second half as the 'target'. He showed that positive lexical adaptation does occur, more strongly for content words than for function words.

Dubey et al. used Church's method to evaluate adaptation for selected syntactic constructions in the Brown and Switchboard corpora (Dubey et al., 2006). They reported positive adaptation for each of the syntactic constructions they considered.

Church's measure was developed to identify the most useful features for information retrieval, rather than for study of adaptation *per se*. Consequently, it has several disadvantages for studying adaptation directly:

- For each feature, this method provides an answer to the question "Did the feature occur in the prime/target?"; however, it does not take into account the frequency of occurrence of a feature, so cannot be used to measure the strength of adaptation
- This method cannot be used to identify adaptation in a single document or between a pair of documents
- This method under-reports adaptation in frequently occurring features

In recent work, Reitter *et al.* (Reitter et al., 2006) investigated syntactic adaptation in Switchboard and Maptask. Instead of using Church's method, they used logistic regression to examine short-term priming effects within a small window of time in single dialogs. This method permits study of the time course of adaptation, but because it applies within a single document only it does not permit examination of the source of adaptation (recency/partner model).

3 Our Measures

We propose two measures. The first one measures the prevalence of adaptation between two documents, while the second one measures the strength of adaptation.

Throughout this discussion, we will use the term 'document' to refer to a dialog or part of a dialog, and the term 'feature' to refer to any phenomenon (lexical, syntactic, referring expression, dialog act, etc.) that occurs in or is labeled in documents.

To measure the degree to which a feature f exhibits adaptation, we divide the corpus into a collection of 'prime' documents and 'target' documents. For each feature f, we compute the frequency of occurrence of the feature in the 'prime' document (*p*), the 'target' document (*t*), and the corpus as a whole (baseline, or *b*). One may use relative frequencies rather than absolute frequencies, or smooth low-frequency features; we do not do this in the experiments reported in this paper because earlier experiments showed that these did not change our results. Both of our measures compare *p* and *t* to *b*. We use the notation $f \in D$ as a shortcut to indicate that the frequency of occurrence of f in document D is greater than the baseline frequency for f.

3.1 Measure 1: Adaptation Ratio

This measure is a modification of Church's measure in two ways. First, it uses the frequency of occurrence of each feature in each document rather than merely its presence or absence. Second, instead of using Church's prior we use an estimate of the probability of feature co-occurence in prime and target by *chance*.

Chance The probability of a feature co-occuring in prime and target by chance is the product of probabilities of its occurence in prime and target independently, assuming independence of the two.

$$P(f \in prime \cap f \in target) =$$

$$P(f \in prime) * P(f \in target) \quad (1)$$

For N $(prime, target)$ dialog pairs where feature f occurs more that *b* times in P $primes$ and more than *b* times in T $targets$, the probability of chance co-occurrence of f in $prime$ and $target$ can be approximated by:

$$chance = (P/N) * (T/N) \quad (2)$$

+Adapt Church defines positive adaptation for a feature f as follows:

$$+adapt = Pr(f \in target \mid f \in Prime) \quad (3)$$

which we approximate as:

$$+adapt = T \cap P/P \quad (4)$$

For this method, we compute for each feature both $chance$ and $+adapt$. We define the *adaptation ratio* as $+adapt/chance$. We sort the features in decreasing order by adaptation ratio. Those at the top of the list exhibit more positive adaptation. We also compute χ^2 to identify features for which the adaptation ratio is significant.

3.2 Measure 2: Adaptation Strength

For this measure, instead of using binary values for each feature indicating presence or absence of that feature in a document, we use the actual frequency of occurrence of the feature in the document.

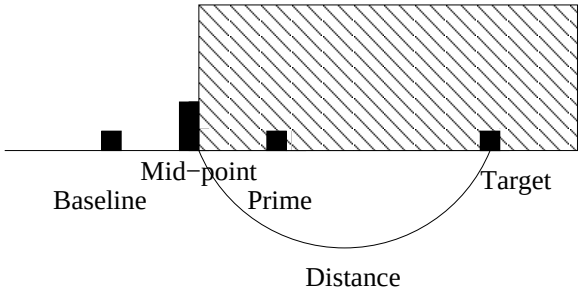

Figure 1: Graphical depiction of Distance

To measure the strength of adaptation on a per-feature basis, we use a $distance$ measure. For a feature f with frequency in $prime$ of p, frequency in $target$ of t and baseline frequency b,

$$distance = t - (p - b)/2 \quad (5)$$

Distance is computed for each feature for each dialog pair. Its value suggests the strength of adaptation for this feature in this dialog pair. Imagine adaptation as a force pulling t towards p and away from b. If there is positive adaptation, then t will be closer to p than to b, as illustrated in Figure 1 (we conservatively chose the midpoint between b and p; a point closer to b could be chosen for a more liberal interpretation of adaptation). We consider a feature to be *adapted* in a pair of dialogs if the target point lies to the right of mid-point in figure 1. We define the *adaptation strength* for a dialog as the average $distance$ over all *adapted* features.

4 Data

The Maptask corpus (Anderson and et. al., 1991) contains 32 sequences of dialogs involving four speakers who discuss routes displayed on maps and trade dialog partners as shown in Table 1. In each dialog, one partner is a *giver* of the route description and the other is a *receiver*. From each dialog sequence, we extract the dialog triples (1,4,6) and (2,3,5). The *follower*, **A**, in the first dialog in each triple (1 or 2) is the *giver* in the second and third dialogs; in the second dialog, **A** speaks with a new conversational partner and in the third dialog **A** speaks with the giver from the first dialog. We hypothesize that persistent recency adaptation will display between the first (*prime*) and second

(*recency*) dialogs in each triple (which are consecutive dialogs for **A**), and partner adaptation between the first (*prime*) and third (*partner*) dialogs in each triple.

Table 2 shows examples of two stem/POS features. *you/DET* occurs in 13 prime dialogs and 11 target dialogs. For 8 (prime, target) dialog pairs it occurs in both dialogs in the pair. For this feature, $chance$ is .14 and $+adapt$ is .62, so we say that this feature exhibits positive adaptation. For the feature *finish/VB* $+adapt$ is less than $chance$, so this feature does not exhibit adaptation.

dlg #	giver	follower	pair1	pair2
1	a1	b1	*prime*	
2	b2	a2		*prime*
3	a2	a1		*recency*
4	b1	b2	*recency*	
5	a2	b2		*partner*
6	b1	a1	*partner*	
7	a1	a2		
8	b2	b1		

Table 1: Maptask dialog order

5 Experiments

In these experiments we ask the following questions:

1. Can we identify the features that affect partner adaptation and recency adaptation?
2. Is partner adaptation or recency adaptation more prevalent?
3. Does the feature frequency in prime affect adaptation of the feature?

We consider two feature types: lexical (word stems, part-of-speech tagged to help distinguish between word senses; and bigrams); and syntactic (productions from the Maptask parse tree annotations).

5.1 Identifying features that exhibit adaptation

In this experiment we identify features with high *adaptation ratios*, looking at both partner and recency adaptation dialog pairs. To minimize noise from infrequently occurring features, in this experiment we only consider features occurring in more than 30% of prime dialogs with frequency higher than the baseline.

	partner	recency
ADJ	right-hand	bottom, right-hand
ADV	when, diagonal	right, well, about
AUX		have
CONJ	if	till, that, so
DET	you, across, on, what, that	my, i, just, that
INTJ	sorri, er,	uh
NOUN	bottom	map
PREP	across, through, along, from	from, by, to
VERB	know, got, take, pass	say

Table 3: Stem/POS features where *adaptation ratio*>1

partner	recency
your left, right-hand side, come to, you come, about the, when you, go round, and round, you got, if you, up toward, a wee, you just, round the, right you, just abov, abov the	no no, my map, okay and, you just, on my, down about, yeah i, you got, down to, have a, i mean, 'til you, just below, just to, now you, no you

Table 4: Bigrams of Stem/POS features where *adaptation ratio*>1

Tables 3, 4, and 5 show the stem/POS, bigram, and syntactic features with *adaptation ratio*>1 and significant χ^2. We observe two interesting categories of features that adapt: perspective and directionality.

In Maptask, speakers can take up a "map-based" perspective (and use words like *north, south, east, west*) or a "paper-based" perspective (and use words like *right, left, top, bottom*). Lexical features indicating perspective are adapted in both partner and recency dialog pairs; the same is true for bigram features. (Other features in this category (e.g. *left, top*) also show adaptation, but occur too infrequently for the adaptation to be significant.)

feature	prime	target	prime ∩target	+adapt	chance
you/DET	13	11	8	0.62	0.14
finish/VB	11	9	1	.09	.10

Table 2: Example lexical features

	partner	recency
advp->		advp
np->	at at ap nn	ap nn; np ap nn; at nn nn; np; np np; pn; ppg nn
pp->	in; rp	pp not pp; ql rp pp; rp aff
s ->	s aff aff s; hv np vp; np; np bez; s s	aff s; np; np s
vp->	vp be np; bez pp; to vp; vb np pp; vb vb pp; vbg pp	advp vp; ber vp; md vp; vb np; vbg; vbg pp vbn pp; vp vp

Table 5: Syntactic features where *adaptation ratio*>1

Directionality in Maptask is indicated by prepositions such as *across, through, along, around* and verbs such as *go* (vs. *take, send*). These prepositions are adapted in both partner and recency dialog pairs, for both lexical and bigram features; the verbs exhibit partner adaptation.

More syntactic features exhibit recency adaptation than partner adaptation.

Table 7 shows adaptation ratio and adaptation strength for some of the syntactic features that were examined in (Dubey et al., 2006). All but the first and last features show comparable partner and recency adaptation ratios. The adaptation strength for the feature $NP->NPPP$ shows stronger partner adaptation than recency adaptation. By contrast, the feature $NP->NN$ shows stronger recency adaptation.

5.2 Comparing partner and recency adaptation

In this experiment, we use *adaptation ratio* and *adaptation strength* to compare partner and recency adaptation. Table 8 shows *adaptation ratio* and *adaptation strength* averaged over all features for each feature type (Stem/POS, Stem/bigram, Syntactic). Positive adaptation for recency dialog pairs in this corpus appears significantly stronger for each feature type, however the probability of chance co-occurrence is also significantly stronger for recency. This explains why there is no significant difference in *adaptation ratio* for lexical features between partner and recency adaptation dialog pairs.

According to the *adaptation ratio* measure, lexical features do not exhibit significant differences between partner adaptation and recency adaptation. However, according to the *adaptation strength* measure, lexical features have stronger adaptation in the partner adaptation dialog pairs. Syntactic features, taken as a whole, do exhibit significantly greater *adaptation ratios* for partner adaptation than for recency adaptation.

Table 9 reports the same measures as Table 8 over the subset of features from Tables 3, 4, 5. The results on the subset of features that exhibit significant positive adaptation are similar to the results for all features.

5.3 Measuring effect of priming frequency on adaptation

This section describes how *adaptation ratio* and *adaptation strength* depend on the frequency of a feature in the prime dialog. Table 10 shows the average *adaptation ratio* and *adaptation strength* values for varying thresholds on the prime: $prime > baseline, prime > baseline + 1, prime > baseline + 2$. The *adaptation ratio* does not depend on variations in the prime dialog frequencies; however, *adaptation strenth* increases as the thresh-

feature	Adapt. ratio		Adapt. strength	
	partner	recency	partner	recency
across	7.314	4.655	0.285	3.452
sorri	4.180	1.741	0.410	0.161
through	5.642	3.385	0.785	1.285
i	1.714	3.0	7.240	8.573
uh	3.413	5.973	1.054	0.471
sai	1.693	5.642	2.430	4.680
about the	4.478	1.492	0.640	2.016
right-hand side	5.924	3.022	2.099	1.640
when you	5.642	2.987	0.660	0.493
my map	2.418	7.052	1.816	0.416
on my	3.173	6.770	1.328	0.328
to be	0.846	3.847	0.265	1.065

Table 6: Comparison of *adaptation ratio* and *adaptation strength* between partner and recency adaptation dialog pairs for the features that have highest differences between the ratios

feature	Adapt. ratio		Adapt. strength	
	partner	recency	partner	recency
NP->NP PP	1.896	2.6	31.699	17.249
NP->NN	2.963	2.963	0.781	2.656
NP->DT NN	3.048	3.048	0.445	0.695
NP->DT AP NN	2.308	3.077	0.254	0.503

Table 7: Adaptation to chance ratio and adaptation strength for the syntactic features examined by Dubey.

old for the prime dialog increases for both recency and user-primed dialog pairs. This trend illustrates that higher occurrence of a feature in the prime dialog causes stronger adaptation (higher frequency of a feature in target), but has no effect on the probability of adaptation.

6 Conclusion

In this paper, we presented two methods for measuring adaptation in dialog. Our *adaptation ratio* measure, a variation on Church's measure of adaptation, evaluates how likely a feature is to appear in a target document with frequency $>$ average if it appears in the prime document with frequency $>$ average. Our *adaptation strength* measure evaluates the strength of adaptation. These measures have several advantages over those used in previous work. Comparing the frequency to average instead of using a binary 'occurred'/'did not occur' allows us to measure effect on both frequent and infrequent features. We think that our measure of *prior* is more sound for measuring adaptation in a relatively small corpus of dialog pairs. Evaluation of adaptation strength allows us to measure adaptation of a feature in single dialog pair.

We used these measures to compare adaptation in partner- and recency-primed dialog pairs. We showed through a series of experiments using the Maptask corpus that these measures can identify features that exhibit variation and can be used across dialogs to evaluate the presence and strength of partner and recency adaptation.

We are still not satisfied with these measures. Some drawbacks to our measures include:

- The *adaptation strength* measure does not take into account the probability of a feature repeating in the same document; some features may be likely to repeat independent of priming.
- In the *adaptation ratio* measure we cut off features that occurred less than 30% in the prime.

feature	Adaptation ratio		Adaptation strength	
	partner	recency	partner	recency
Stem/POS	2.64	2.71	3.46	3.67*
Stem/bigram	2.99	3.03	1.71	1.91*
Syntactic	2.71	2.92*	4.70*	4.11

Table 8: Adaptation ratio and adaptation strength averaged over all features. * indicates significant difference between partner and recency adaptation ($p<.05$)

feature	Pr(+adapt)/Pr(Chance)		Adapt. Strength	
	partner	recency	partner	recency
Stem/POS	3.36	3.15	3.71	3.82
Stem/bigram	3.86	3.68	1.30	1.62*
Syntactic	3.09	3.36*	5.49*	4.99

Table 9: Adaptation ratio and adaptation strength averaged over significant features listed in Tables 3, 4 and 5. * indicates a significant difference between partner and recency adaptation ($p<.05$)

Taking a different cut-off may influence the result.

We hope to address these issues in future work.

In current work, we are incorporating models of adaptation to syntactic and lexical choice into our RavenCalendar dialog system (Stenchikova et al., 2007). We are creating a tight integration between parsing, dialog management and response generation so that words and syntactic constructions used by the user can be highly salient for the system, and ones used by the system are available for interpretation of user utterances (cf. (Isard et al., 2006)). In experiments with this system, we plan to use our adaptation measures to evaluate user adaptation to system behavior for different system adaptation rates.

References

A. Anderson and et. al. 1991. The HCRC map task corpus. *Language and Speech*, 34:351–366.

L. Bell, J. Boye, J. Gustafson, and M. Wiren. 2000. Modality convergence in a multimodal dialogue system. In *Proceedings of GOTALOG*.

S. Brennan and H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology*, 22(6):1482–1493.

S. Brennan. 1996. Lexical entrainment in spontaneous dialog. In *Proceedings of ISSD*, pages 41–44.

P. Brown and G. Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19:441–472.

S. Brennan C. Lockridge. 2002. Addressees' needs influence speakers' early syntactic choices. *Psychonomics Bulletin and Review*. Psych.

T. Chartrand and J. Bargh. 1999. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76:893–910.

K. Church. 2000. Empirical estimates of adaptation: the chance of two noriegas is closer to p/2 than p2. In *Proceedings of the ACL*.

A. Dubey, P. Sturt, and F. Keller. 2006. Parallelism in coordination as an instance of syntactic priming: evidence from corpus-based modeling. In *Proceedings of EMNLP*.

J. Gustafson, A. Larsson, R. Carlson, and K. Hellman. 1997. How do system questions influence lexical choices in user answers? In *Proceedings of Eurospeech*.

W. Horton and R. Gerrig. 2002. Speakers' experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47:589–606.

A. Isard, C. Brockmann, and J. Oberlander. 2006. Individuality and alignment in generated dialogues. In *Proceedings of INLG*.

T. Kraljic and S. Brennan. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50:194–231.

	num	%adapted		avg. adapt. strength	
		partner	recency	partner	recency
p>b	151.7	.14	.17	2.42	2.55
p>b+1	78	.12	.14	3.47	3.59
p>b+2	51.8	.12	.15	3.94	3.82

Table 10: Average distance measures for *adapted* features (Stem/POS only) [1]

R. Kraut, S. Fussell, and J. Siegel. 2003. Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18(1–2).

M. Pickering and S. Garrod. 2004. Toward a mechanistic psychology of dialog. *Behavioral and Brain Sciences*, 27:169–190. Psych.

M. Pickering, H. Branigan, A. Cleland, and A. Stewart. 2000. Activation of syntactic priming during language production. *Journal of Psycholinguistic Research*, 29(2):205–216.

D. Reitter, F. Keller, and J. Moore. 2006. Computational modelling of structural priming in dialogue. In *Proceedings of HLT/NAACL*.

G. Skantze. 2002. Coordination of referring expressions in multimodal human-computer dialogue. In *Proceedings of ICSLP*.

S. Stenchikova, B. Mucha, S. Hoffman, and A. Stent. 2007. RavenCalendar: A multimodal dialog system for managing a personal calendar. In *Proceedings of HLT/NAACL*.