system	$Acc_{message}$ (%)	Acc_{riding} (%)
FRQ	54.82	63.14
MJR	36.48	36.63
INC	53.29	78.03
NGR (uni+bi+tri)	62.02	79.65
JDG	66.23	78.68
Crystal (uni+bi+tri)	**73.07**	**81.68**

Table 4. System performance with accuracy per message ($Acc_{message}$) and accuracy per riding (Acc_{riding}): FRQ, MJR, INC, NGR, JDG, and *Crystal*.

Features	$Acc_{message}$ (%)	
	NGR	***Crystal***
uni	60.49	72.03
bi	58.79	71.81
tri	54.04	69.57
four	47.25	67.64
uni + bi	61.54	72.93
uni + tri	61.36	72.20
uni + four	60.70	72.84
bi + tri	58.68	72.26
bi + four	58.54	72.17
uni + bi + tri	**62.02**	**73.07**
uni + bi + four	61.75	72.30
uni + tri + four	61.34	72.30
bi + tri + four	58.42	72.62
uni + bi + tri + four	61.96	73.01

Table 5. System performance with different features: Pure n-gram (NGR) and Generalized n-gram *Crystal.*

higher than FRQ and INC. The best accuracy of our system was also obtained with the combination of unigram, bigram, and trigram features.

The JDG system, which uses positive and negative sentiment word features, had 66.23% accuracy. This is about 7% lower than *Crystal*. Since the lower performance of JDG might be related to the number of features it uses, we also experimented with the reduced number of features of *Crystal* based on the *tfidf* scores[11]. With the same number of features (i.e., 1635), *Crystal* performed 70.62% which is 4.4% higher than JDG. An interesting finding was that NGR with 1635 features performed only 54.60% which is significantly

[11] The total number of all features of *Crystal* is 689,642.

Patterns in *WIN* class	Patterns in *LOSE* class
PARTY_will_win	want_OTHER
PARTY_hold	PARTY_don't_have
PARTY_will_win_this	OTHER_and
PARTY_win	the_PARTY
will_go_PARTY	OTHER_will_win
PARTY_will_take	OTHER_is
PARTY_will_take_this	to_the_OTHER
PARTY_is	and_OTHER
safest_PARTY	results_OTHER
PARTY_has	OTHER_has
go_PARTY_again	to_OTHER

Table 6. Examples of frequent features in WIN and LOSE classes.

lower than both systems. This indicates that the 1635 pure n-gram features are not as good as the same number of sentiment words carefully chosen from a dictionary but the generalized features of *Crystal* represent the predictive opinions better than JDG features.

Table 5 illustrates the comparison of NGR (without feature generalization) and *Crystal* (with feature generalization) in different feature combinations. *uni, bi, tri,* and *four* correspond to *unigram, bigram, trigram*, and *fourgram*. Our proposed technique *Crystal* performed always better than the pure n-gram system (NGR). Both systems performed best (62.02% and 73.07%) with the combination of unigram, bigram, and trigram (uni+bi+tri). The second best scores (61.96% and 73.01%) are achieved with the combinations of all grams (uni+bi+tri+four) in both systems. Using fourgrams alone performed worst since the system overfitted to the training examples.

Table 6 presents several examples of frequent n-gram features in both WIN and LOSE classes. As shown in Table 6, lexical patterns in the WIN class express optimistic sentiments about PARTY (e.g., PARTY_will_win and go_ PARTY_again) whereas patterns in the LOSE class express pessimistic sentiments (e.g., PARTY_don't_have) and optimistic ones about OTHER (e.g., want_OTHER).

Evaluation2: In this evaluation, we use Acc_{riding} computed as the number of ridings that a system correctly predicted, divided by the total number of ridings. For each riding *R*, systems pick a party that obtains the majority prediction votes from messages in *R* as the winning party of *R*. For ex-