IJCLA VOL. 5, NO. 1, JAN-JUN 2014, PP. 73–86
RECEIVED 06/01/14 ACCEPTED 06/02/14 FINAL 05/04/14

Linguistic Features Predict the Truthfulness of Short Political Statements

VIVEK V. DATLA,[1] KING-IP LIN,[1] AND MAX M. LOUWERSE[2]

[1] *University of Memphis, USA*
[2] *Tilburg University, The Netherlands*

ABSTRACT

Checking the truth value of political statements is difficult. Fact checking computationally has therefore not been very successful. An alternative to checking the truth value of a statement is to not consider the facts that are stated, but the way the statement is expressed. Using linguistic features from seven computational linguistic algorithms, we investigated whether truth-false statements and the definitiveness with which the statement is expressed can be predicted using linguistic features. In a training set we found that both distinctiveness and truthfulness of the statement predicted linguistic variables. These variables corresponded to those mentioned in deception literature. Next, we used a new set of political statements and determined whether the same linguistic variables would be able to predict the definitiveness and truthfulness of the statement. Given the fact that the political statements are short, one-sentence statements, allowing for a large variability in linguistic variables, discriminant analyses showed that the function obtained from the training set allowed for an accurate classification of $57 - 59\%$ *of the data. These findings are encouraging, for instance for first analysis on the truth value and verifiability of political statements.*

1 INTRODUCTION

In a speech, Rick Santorum, runner for the Republican presidential nomination, 2011, said: *"[T]hey have voluntary euthanasia in the Netherlands, but half the people who are euthanized every year, and it's 10 percent of all deaths for the Netherlands, half of those people are euthanized* *involuntarily at hospitals because they are older and sick."* Political statements like these might sound convincing and definite. Yet, it is unclear whether these statements are actually true. Finding out whether they are takes a considerable amount of investigative work.

One can investigate the number of deaths and euthanizations in the Netherlands and conclude that the statement is false. Such a task can perhaps be performed computationally, whereby a computational algorithm interprets a statement, tracks down the facts, and compares the truth value of these facts. However, a successful algorithm is not yet on the computational linguistic horizon [1]. An alternative might lie not so much in identifying the truth value of the facts being stated, but in investigating the style a statement is expressed in.

Speakers generally follow guidelines for a smooth conversation, summarized by [2] in four maxims of communication. The maxim of quantity postulates that the speaker should not say more, or less, than what is needed, the maxim of relation postulates the speaker should be relevant to the purposes of the conversation, and the maxim of manner postulates the speaker should be clear and orderly. Importantly for the current paper, the maxim of quality states "do not say what you believe to be false" and "do not say that for which you lack adequate evidence".

Some political statements, like the one by Rick Santorum quoted earlier, happen to be false and lack any evidence. We can investigate whether the style of the statement gives away cues that indicate the speaker is not quite sure that he says what s/he believes to be true and only says that for which s/he has adequate evidence. Such an investigation on determining the definitiveness and the truthfulness of political statements is the topic of this paper.

A politician may use a formal style of language in order to create the impression that s/he presents precise, objective information, while s/he really wants to hide the exact details of his/her policy [3]. If the speakers purposefully violate Grice's maxim of quality, they leave non-linguistic and linguistic footprints in their attempts to hide the truth [4].

There is of course a distinction between not quite telling the truth and actual deceiving. A speaker might not tell the truth because s/he does not have the facts readily available but needs to say something, or because the facts cannot be stated because of political, strategic, or social reasons (maxim of relevance). However, regardless of the motivating behind hiding the truth, the non-linguistic and linguistic footprints in the speaker's attempts to hide the truth might actually be the same in deception and non-truth telling: in both cases the speaker has an increased cognitive load because of not wanting to tell the truth, even though the motivations behind not telling the truth might be different.

There are various studies that report on verbal and non-verbal footprints left behind in deception. For instance, in the context of police interviews people telling a lie used fewer illustrations, had an increase in pauses, and an increase in the latency period, most likely due to the increased cognitive load, i.e., the focus on both not telling the truth and the actual telling the truth [5]. A review of 116 deception studies by [4] showed that lies had more verbal and vocal uncertainty, less verbal and vocal immediacy, were more ambivalent, less plausible and had less logical structure, with less contextual embedding.

DePaulo et al. [4] found that deceptive communication had fewer first-person singular pronouns, fewer third-person pronouns, more negative emotion words (e.g., hate, anger, enemy), fewer exclusive words (e.g., but, except), and more motion verbs (e.g., walk, move, go). [6] investigated statements from speakers who were asked to be deceptive in asynchronous computer-mediated communication (CMC). Participants were asked to write stories on five different topics, with one group of participants asked to not tell the truth. The untrue stories consisted of fewer words, fewer first person pronouns, more questions, and more words pertaining to senses (e.g., see, listen). This finding is consistent with [7] findings.

DePaulo et al. [4] argued that the motivation to not tell the truth plays an important role in the linguistic features of the statements. The settings of typical laboratory experiments lack a participant's motivation. That is, when untrue statements in society are investigated the stakes are higher. In the case of a politician not telling the truth could mean the difference between being considered credible or not, between voted into office or not. It can therefore be expected that the verbal footprints are more easily to detect than when speakers are less motivated to tell or hide the truth.

Much of the literature investigating linguistic cues in statements where the speaker says what he/she believes to be false uses passages or paragraphs. Indeed, if verbal footprints of not telling the truth are left, they will be more obvious when more data from a speaker is available. However, there often is only limited data available. Twitter messages, Facebook comments, or other brief comments do not allow for lengthy text. In addition while the overall message of a conversation may not be false, individual statements within it may be inaccurate or fabricated. Therefore, even though it might be easier to detect deception in large text samples [6, 7], the current study investigated whether linguistic cues from short political statements also predict definitiveness and truthfulness.

We used statements from politifact.com [8], because they consist of recent relatively short—on average approximately 18 words—statements from a variety of politicians that are checked on their accuracy.

Politifact is a project of the Tampa Bay Times, a Florida-based media organization, which won the prestigious Pulitzer Prize for its fact checking during the 2008 presidential election campaigns. Statements include those by members of congress, state legislators, governors, mayors, the president, cabinet secretaries, lobbyists, people who testify before Congress etc. Politifact uses the following categories to represent the truthfulness of a statement:

- True: The statement is accurate
- Mostly true: The statement is accurate but needs clarification
- Half true: The statement is partially accurate
- Mostly false: The statement contains an element of truth but ignores some information
- False: the statement is not accurate
- Pants on fire: The statement is absurdly false

These six categories allow for two pieces of information. First, statements can be categorized in true and false statements. However, we can do this on the basis of a strict criterion (true versus false and pants on fire) or a more lenient criterion (true and mostly true versus false, mostly false and pants on fire). We also did a different analysis by making a distinction in truthfulness (regardless of whether the distinction is made based on a strict or lenient criterion) is the definitiveness of the truth or false value. Half true statements are half true and half false, and therefore are not definitive; it can be expected that stylistic cues give away to what extent the speaker expresses a statement in a more (true/false) or less (half true/false) way; see Table 1.

The Politifact statements and the 2 definitiveness $\times$ 2 truthfulness (strict and lenient) categories allow us to train the linguistic characteristics of each political statement on its respective category. The resulting categorization function from this training can then be tested on a test set of new political statements.

2 Linguistic Features

To train the model, a wide range of computational linguistic dimensions was selected, including syntactic and semantic algorithms. These algo-

Table 1. Overview of the politifact categories

		True	False
Definitive	**Strict**	true	false, pants on fire
	Lenient	true, mostly true	false, mostly false, pants on fire
Indefinitive		half-true	

rithms can, generally, be classified into general structural (e.g., word count), syntactic (e.g., connectives) and semantic (e.g., word choice) dimensions of language. Five different algorithms were used, categorized in Figure 1.

For general linguistic features, we used the frequency of 67 linguistic features used in [9]. These features in primarily operate at the word level (e.g., parts-of-speech) and can be categorized as tense and aspect markers, place and time adverbials, pronouns and pro-verbs, questions, nominal forms, passives, stative forms, subordination features, prepositional phrases, adjectives and adverbs, lexical specificity, lexical classes, modals, specialized verb classes, reduced forms and dispreferred structures, and co-ordinations and negations.

For WordNet [10] 150,000 words in 44 base types were selected, including 25 primitive groups for nouns (e.g. time, location, person etc.), 15 for verbs (e.g. communication, cognition, etc.), 3 groups of adjectives and 1 group of adverbs.

The linguistic category model (LCM) gives insight into the interpersonal language use. The model consists of a classification of interpersonal (transitive) verbs that are used to describe actions or psychological states and adjectives that are employed to characterize persons. In order to capture the various emotions expressed by the statement we have used the emotion words given by [11], classified into two classes broadly basic emotions (anger, fear, disgust, happiness etc.) and complex emotions (guilt, pity, tenderness etc.). The basic emotions indicate no cognitive load hence they are also called as raw emotions, whereas the complex emotions indicate cognitive load.

Interclausal relationships were captured using [12] parameterization, including positive additive, (also, moreover), negative additive (however, but), positive temporal (after, before), negative temporal (until), and causal (because, so) connectives. In order to get the frequencies of the words we have used CELEX database [13]. The CELEX database consists of 17.9

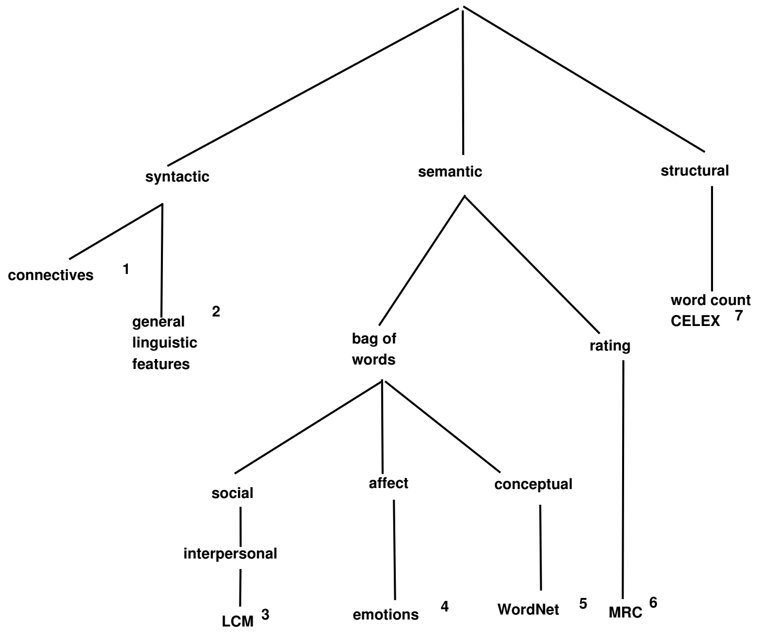

Fig. 1. Overview of computational linguistic algorithms used. [1]Louwerse (2002), [2]Biber (1988), [3]Semin & Fiedler (1991), [4]Johnson-Laird & Oatley (1989), [5]Miller et al. (1990), [6]Coltheart (1981), [7] Baayen, Piepenbrock, & Gulikers (1995)

million words taken from both spoken (news wire and telephonic conversations) and written (newspapers and books) corpora.

In addition, we used the MRC Psycholinguistic Database [14], to get linguistic measures such as familiarity, concreteness, imaginability and meaningfulness. For each political statement collected from Politifact.com [8] we processed the features for the 7 computational linguistic algorithm, normalized for the number of words per statements, and the scores were treated as a vector.

3 Training

A total of 1576 political statement were downloaded from Politifact.com sentences) as training data. These political statements came from April, 2012. The break down of the various categories for the training data are as follows: 21% true, 19% mostly true, 22% half true, 15% mostly false, and 23% false.

4 RESULTS AND DISCUSSION

A mixed effects regression model was run on each of the linguistic features with the category as independent variable and individual speaker as a random factor, to avoid any speaker bias [13]. The model was fitted using the restricted maximum likelihood estimation (REML) for the dependent. F-test denominator degrees of freedom were estimated using the Kenward-Roger's degrees of freedom adjustment to reduce the chances of Type I error [15].

We predicted that those patterns found in the deception studies discussed earlier would be found in the computational linguistic scores.

4.1 *Truthfulness*

As the results of the mixed effect regression model in Table 3 shows, truthfulness explained the variance of 20 linguistic variables, with similar patterns and variables for the strict and lenient categories.

The results show that various verb categories (cognitive, communicative, modals, predicated modals) explain the difference in truthfulness, a finding in line with the idea that these verbs increase verbal immediacy and cognitive load. The results in Table 3 are also in line with [4, 6, 7, 16, 17] who have all shown that negative emotions are related to deception, in our analysis emotions came to significance while classifying between true and false in the lenient case.

To put the findings reported in Table 3 in perspective, we linked each of the findings to a corresponding finding in the deception literature using the studies reported in Table 2.

4.2 *Definitiveness*

As the results of the mixed effect regression model in Table 4 shows, definitiveness explained the variance of 20 linguistic variables. Importantly, the direction of the significant linguistic features is similar across the strict and lenient categories. In both strict and lenient cases, variables such as concreteness, word count, variety in the tokens in the statement, positive connectives, has shown up to be significant. The results indicate that if the statement is more concrete or has high imagery score then

Table 2. Relevant Deception Studies. The number listed in the first column corresponds to the number used in the first column of the results.

ref	Literature
1	Newman et al., (2003) [7]
2	Tausczik & Pennebaker, (2010) [16]
3	Hancock, et al, (2007) [6]
4	DePaulo et al., (2003) [4]
5	Toma & Hancock (2010) [17]
6	Louwerse, et al., (2010) [18]

Table 3. Variables that explain truthfulness of a political statements. First columns gives references to deception literature. Superscript in the second column gives reference to the computational linguistic model. Last columns give the t-values to show the direction of the effect (**: $p \leq 0.01$, *: $0.01 \leq p \leq 0.05$).

		Condition	
Ref	Language Features	Strict	Lenient
1, 3, 4 ,5	Positive connectives[1]	-2.13*	-3.04**
1, 2, 3, 4, 5	Caused Emotions[4]	1.07	2.01*
1 ,2	Social verbs[5]	1.59	2.90**
1, 2, 3, 4, 5	Emotion verbs[5]	1.17	2.93**
1, 2, 3, 4, 5	Cognitive verbs[5]	1.69	3.56**
1, 2	Communication verbs[5]	2.40*	4.18**
1	Possession verbs[5]	1.99*	3.03**
2	Prepositions[2]	-2.6**	-3.49**
	Second person pronouns[2]	2.59*	3.29**
1	Modal verbs[2]	1.74	3.8**
	Numbers[2]	-2.02*	-2.99**
	CELEX frequency[7]	-2.57*	-3.2**
1, 3, 4 ,5	Temporal positive connectives[1]	-1.14	-3.24**
1, 3, 4 ,5	Additive positive connectives[1]	-2.17*	-2.9**
1, 3, 4 ,5	Temporal connectives[1]	-1.16	-2.81**
	Private verbs[2]	2.34*	2.04*
1	Predicated Modality[2]	1.50	3.05**
	Emphatics[1]	-2.17*	-1.71
3,5	Brown Frequency[6]	2.34*	2.1*

it is more likely to be an indefinite(half true) statement, than a definite (true/false) statement. This corresponds with the literature. [16] indicate pronoun use, emotionally toned words, and prepositions and conjunctions that signal cognitive load are linked to behavioral and emotional

Table 4. Language features that help in detecting definitiveness of the truth value in the statement. First columns gives references to deception literature. Superscript in the second column gives reference to the computational linguistic model. Last columns give the t-values to show the direction of the effect (**: $p \leq 0.01$, *: $p \leq 0.05$).

		Condition	
Ref	Language Features	Strict	Lenient
3, 5	Word Count[7]	-3.96**	-3.79**
4	Token types[7]	3.23*	3.6*
	Concreteness with type[6]	-2.08*	-2.3*
	Concreteness with token[6]	-2.71*	-2.7*
1, 3, 4, 5	Positive connectives[1]	-2.64*	-2.4*
1, 3, 4 ,5	Additive positive connectives[1]	-2.30*	-2*
1, 3, 4, 5	Additives[1]	-2.54*	-1.85
	Consumption Verbs[5]	-2.20*	-2.04*
1	Communication Verbs[5]	-2.08*	-2.2*
1, 2, 3, 4, 5,6	State Action Verbs[3]	-2.49*	-1.57
	Public Verbs[2]	2.48*	2*
1	Prepositions[2]	-3.66**	-2.9**
1	Auxiliary Verbs[2]	2.54**	1.64

outcomes. Similarly, [7] indicate that self-references, negative emotion words and cognitive complexity play an important role when people try to deceive. In our analysis we find that connectives help in classifying between definite and indefinite sentences, with higher frequencies of connectives yielding more complex sentences and consequently higher cognitive load.

5 TESTING

A total of 1597 political statements from January 2013 were downloaded as a test set. The breakdown of the various categories for the test data are as follows: 14% true, 30.8% false, 16.7% mostly-true, 15.2% mostly-false and 22.4% half-true statements.

As the sizes of the categories of statements are not equal, this makes the discriminant analysis classify all the instances of the classes to the post popular class in data. In order to make sure that we have not made a special selection of statements that make the two classes, we conducted 1000 Monte Carlo simulations on both truthfulness and definiteness cases, and also in their strict and lenient sub cases, to pick two equal classes, for discriminant analysis. This also helps to improve the robustness of our classification.

In order to make sure that we do not overfit the model with all the variables that came to significance in the mixed effect regression model, we selected a random set from the 1000 sets created by the Monte Carlo simulations from the strict case of both, True Vs. False, and Definite (true/false) Vs. Indefinite(half true) classification. Then selected the smallest set of variables in each case that classify them into their respective classes. In case of True Vs. False, we got the best result for the classification by taking the variables, social verbs, modal verbs, numbers, private verbs and brown frequency.

In the case of classification between definite and Indefinite statements, we got the best classification on the random set, by using the variables word count, prepositions, token types, concreteness with token types, positive connectives and additives.

We have used the same variables for the classification in the strict and the lenient case. The results of the classification are in shown in the Table 5 and Table 6. In case of classifying between the True and False statements, communication verbs (announce, argue, express etc.), social verbs (observe, upgrade, permit etc.) and modal verbs (can, could, may etc.) which indicate cognitive load and verbal immediacy were significant. These categories of words are also referred in the deception literature [7]; [16], indicating that even in constricted context these categories help in classifying true and false statements.

In case of classifying between the definite and indefinite statements in both strict and lenient case, we are able to classify significantly between the cases with accuracy of about 58% on an average over 1000 runs. The classification is more significant in the strict case compared to the lenient case. Table 5 shows the results averaged over 1000 runs for strict and lenient case, which we obtained for classifying the statements into true and false. The results indicate that we need more context in classifying true and false statements, as in the lenient case we are able to classify between the true and false cases more significantly.

Table 6 indicates the accuracy for classifying between definite and indefinite statements in strict and lenient cases. Even though the accuracy on average over 1000 runs is only 59% given that chance is 50%, it is a significant result as we are analyzing the statements with very few words. The classification in the strict case and the lenient cases the classification are significant. The significance of classification is smaller in the lenient

Table 5. Classification between true and false statements

Strict	True	False	Overall
True	42.6%	57.4%	56.8%
False	28.9%	71.1%	
$\chi^2(1, N = 5) = 16.89, p < 0.024$			
Lenient	True	False	Overall
True	44.7%	55.3%	55.9%
False	32.8%	67.2%	
$\chi^2(1, N = 5) = 23.67, p < 0.0041$			

Table 6. Classification between Definite and Indefinite Statements

Strict	True	False	Overall
True	59.2%	41.8%	57.9%
False	43.5%	56.5%	
$\chi^2(1, N = 6) = 27.54, p < 0.002$			
Lenient	True	False	Overall
True	56.6%	43.4%	55.9%
False	44%	56%	
$\chi^2(1, N = 6) = 17.4, p < 0.04$			

case, this is due to the fact the mostly-true, half-true, and mostly false are contiguous on the deception scale.

6 Conclusion and Future Work

This study investigated whether linguistic features can be predicting the truth value and the definitiveness of the truth value in short political statements. Using a training set of one-sentence political statements, we investigated whether linguistic features obtained from seven computational linguistic algorithms across syntactic, semantic and structural dimensions showed a relationship with truthfulness and definitiveness. We thereby used a strict criterion and a more lenient criterion. Results showed that a similar set of linguistic variables explained these categories. In a testing set of a new set of political statements we then tested whether the same variables explained the truthfulness and definitiveness categories. The results showed they did, supporting the conclusion that linguistic features can help determining to what extent political statements are true and to what extent this decision can be made with certainty.

Interestingly, the linguistic variables that have been identified as predictors of truthfulness and definitiveness match the variables that have been identified as linguistic cues to deception. A large body of literature has investigated whether deceivers leave linguistic footprints in their deception. However, rather than with the purpose of deceiving, we assume that the speakers of the short political statements of the current study had valid reasons to not quite tell the truth. The findings reported here might not overwhelm. A 55-60% discrimination score is not high. Yet, the fact that such a score is significant, that the variables behind the score are consistent across training and testing, and that this score is obtained with a small language unit (about one sentence), makes the findings reported in the current study remarkable nonetheless.

The computational linguistic means of predicting truthfulness and definitiveness should certainly not stand on their own in evaluating short political statements. However, they can fulfill a supporting role. Computational linguistic algorithms such as the ones discussed can identify whether statements can be easily checked and whether there is an initial likelihood that supporting evidence can or cannot be found. In the day and age of Twitter and Facebook with many short statements, having a tool that filters whether a statement is the truth and nothing but the truth or not, might be very welcome.

REFERENCES

1. Mihalcea, R., Strapparava, C.: The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. ACLShort '09, Stroudsburg, PA, USA, Association for Computational Linguistics (2009) 309–312
2. Grice, H.P.: Logic and conversation. Syntax and Semantics **3** (1975) 41–58
3. Heylighen, F., Dewaele, J.M.: Variation in the contextuality of language: An empirical measure. Foundations of Science **7**(3) (2002) 293–340
4. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological bulletin **129**(1) (2003) 74
5. Vrij, A., Edward, K., Roberts, K., Bull, R.: Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal Behavior **24**(4) (2000) 239–263
6. Hancock, J.T., Curry, L.E., Goorha, S., Woodworth, M.: On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. Discourse Processes **45**(1) (2007) 1–23
7. Newman, M.L., Pennebaker, J.W., Berry, D.S., Richards, J.M.: Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin **29**(5) (2003) 665–675

8. TampaBay Times: Politifact.com. Published online (April 2012)
9. Biber, D.: Variation across speech and writing. Cambridge University Press (1991)
10. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: Five papers on WordNet. In Fellbaum, C., ed.: WordNet: An Electronic Lexical Database. MIT Press (May 1998)
11. Johnson-laird, P.N., Oatley, K.: The language of emotions: An analysis of a semantic field. Cognition and Emotion **3**(2) (1989) 81–123
12. Louwerse, M.: An analytic and cognitive parametrization of coherence relations. Cognitive Linguistics **12**(3) (2001) 291–316
13. Baayen, H.R., Piepenbrock, R., Gulikers, L.: The CELEX lexical database. Release 2 (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, Pennsylvania (1995)
14. Coltheart, M.: The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology **33**(4) (1981) 497–505
15. Littell, R., Stroup, W., Freund, R.: SAS for Linear Models,: Fourth Edition. Wiley series in probability and statistics. SAS Institute (2002)
16. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology **29**(1) (2010) 24–54
17. Toma, C.L., Hancock, J.T.: Reading between the lines: Linguistic cues to deception in online dating profiles. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. CSCW '10, New York, NY, USA, ACM (2010) 5–8
18. Louwerse, M.M., Lin, K., Drescher, A., Semin, G.: Linguistic cues predict fraudulent events in a corporate social network. In Ohlsson, S., Catrambone, R., eds.: Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, TX, The organization, Cognitive Science Society (2010) 961–966

VIVEK V. DATLA
DEPTARTMENT OF COMPUTER SCIENCE,
UNIVERSITY OF MEMPHIS,
TN 38152, USA
E-MAIL: <VVDATLA@MEMPHIS.EDU>

KING-IP LIN
INSTITUTE FOR INTELLIGENT SYSTEMS,
UNIVERSITY OF MEMPHIS,
365 INNOVATION DRIVE, MEMPHIS, TN 38152, USA
E-MAIL: <DAVIDLIN@MEMPHIS.EDU>

MAX M. LOUWERSE
TILBURG CENTRE FOR COGNITION AND COMMUNICATION (TICC),
TILBURG UNIVERSITY,
PO BOX 90153, 5000 LE, TILBURG, THE NETHERLANDS
E-MAIL: <MLOUWERSE@MEMPHIS.EDU>