Language Patterns in Text Messages of Electronic Negotiations: A Preliminary Study*

Marina Sokolova, Stan Szpakowicz and Vivi Nastase

School of Information Technology and Engineering
University of Ottawa, Ottawa, Canada

Abstract

Negotiation support systems often allow an exchange of messages that help explain better the offers and positions of the negotiators. Collections of such messages can be analyzed using Natural Language Processing techniques. We work with a large collection that the Inspire system accumulated in several years of use as a tool for teaching negotiation. Messages are unedited and mostly very noisy. Since the topic of negotiations mediated by Inspire is constant, the message texts can be treated as belonging to a closed domain.

We introduce a procedure that uses the Inspire text data to classify negotiations as successful or failed, and to find language patterns characteristic of these two classes. Our procedure can apply to any similar collection of texts that accompany electronic negotiation or other comparable processes mediated by Web-based systems. The results show that, even in the early stages of negotiation, certain patterns in the language of the messages do predict whether the negotiation will succeed.

*This is a submitted version of a paper that will appear, in the final form, in the proceedings of the Seventeenth Canadian Conference on Artificial Intelligence, London, Ontario, May 2004.

1 Introduction

We investigate how language reflects success or failure of negotiations. Our purpose is to build a language model of texts that accompany fixed-problem electronic negotiations (e-negotiations) performed with the help of a negotiation support system (NSS). A language model, depending on its parameters, can represent a successful or failed negotiation. Models relate success or failure of negotiations to such text characteristics as word statistics and lexical semantics.

People proceed in e-negotiations differently than in face-to-face or phone negotiations. They do not have visual or acoustic information to evaluate the process of negotiations and plan their future actions, nor can they use visual or acoustic means to influence the process in order to achieve their goals. Only texts of messages are available to the negotiators. This implies that the language of texts is vitally important for understanding the process and outcome of negotiations. We note that nobody seems to have applied natural language processing (NLP) techniques to the study of the data of e-negotiations, although the Management Science and Artificial Intelligence communities [1, 2, 10, 13, 17, 18] actively investigate the process and data of e-negotiation.

With negotiation analysis integrating decision analysis and game theory [5], language models of e-negotiations are useful for understanding how the negotiators' behaviour reflects their strategic goals and tactical intentions, for balancing the subjective self-evaluation of negotiators by providing the external, and presumably objective, source of evaluation.

In practice language models help design, develop and implement NLP systems capable of handling text data obtained through electronic means (Web data). Web data are characterized by a large number of spelling and grammatical errors, and the uncontrolled use of informal and slang expressions. This makes Web data noisy. The excessive quantity of noise distinguishes Web data from collections of texts communicated through more traditional channels, in particular well edited texts of books, articles and manuals. These characteristics suggest that an NLP system should adjust to the different types of noise. A lexicon-building part of such a system was introduced in [16].

Continuing previous research, we present a new procedure that uses negotiation text data to classify successful and failed negotiations. We employ it to the finding of language patterns characteristic of successful or failed negotiations. To test the procedure, we have applied it to a sample of data from the NSS Inspire [4] and compared the outcome with the results of a data mining exercise [7]. The results show that the procedure effectively identifies success in negotiations. This procedure does not depend on the NSS and can work for Web data different from those collected by Inspire.

In section 2 we describe the Inspire data and discuss the challenges they pose. Section 3 presents a classification procedure and reports its results. Section 4 presents a procedure of finding language patterns characteristic of success or failure of negotiations, and reports the results of the procedure on the Inspire data. Section 5 contains conclusions and suggestions for future work.

2 The Inspire Collection

We experiment with a collection of text messages exchanged by negotiators through the Inspire NSS [6, 4]. Inspire is a teaching tool used in the business and management university programs in a number of countries. It provides a medium for exchanging information during the negotiation process. This information includes tables that represent offers and messages that either complement offers or are exchanged between offers.

We have access to the transcripts of 1482 negotiations conducted using Inspire. Each negotiation involves a buyer, who represents Cypress Cycles, and a seller, who represents Itex Manufacturing. These two people seek agreement on buying and selling bicycle parts. One person participates in only one negotiation, so our collection of messages will have well over 2000 authors. While the negotiators' educational and cultural background differs, they have common characteristics: they are university or college students or professionals, they all speak English (albeit many of them as a second language) and they were all given the same manuals and instructions about the negotiation process.

We have collected 14085 messages from the 1482 transcripts. This resulted in a corpus of 827209 word tokens, which correspond to 20990 types. This data presents us with some interesting challenges. Although English was suggested as the language of negotiation, some messages contain words from, or are even entirely written in, other languages (German, Spanish or Russian transliterated in Latin). We must filter out messages not written in English. Other challenges arise from the fact that for most negotiators English is not the first language, so spelling and language mistakes are very common. Since the message exchange is quite informal, editing the text does not seem to have been a concern.

We need to address these problems in order to process the Inspire data with the purpose of future building a language model of successful negotiations. To filter out messages not written in English, we do not focus, as one might expect, on function words, the most common and easiest to detect in well edited texts. The reason is that spelling errors can mask these words ("and" in an English text appears sometimes as "und", erroneously suggesting that the message could be in German). We consider longer words. We build a list of words in foreign languages that are more frequent across messages, then find messages containing those words. If a message contains mostly non-English words, we delete it from the data.

Editing the messages is a bigger problem. Manual analysis of the collected messages has shown four main causes of noise in data:

- words containing non-letter characters;
- use of foreign words within messages written in English;
- use of informal and slang expressions;
- spelling errors, missing punctuation, spaces missing between words, incorrect capitalization.

We semi-automatically solve this problem by using the `ispell` spell checker in Unix, and frequency counts for unigrams in the text (we pick the correct word suggested by `ispell` that appears most frequently in the data).

After filtering and editing the data, we focus on individual words. In order to use them to build a model of the language that accompanies negotiations, we need to group the words by the role they play in the negotiation process. Some words are generic with respect to this task – not technical – while others are specific either to the negotiation process (*offer*, *counter-offer*, and so on) or to the items whose sale is negotiated (in our case, bicycle parts). According to this criterion we identify six semantic zones [16] to which the words in our collection belong: business in general, negotiation processes, communication, bicycle parts, non-technical vocabulary, function words.

The description of the Inspire collection makes it clear that its contributors have similar educational characteristics and discuss a fixed topic. We want to know how the conditions of educational similarity and topic pre-assignment affect the growth of the language. The Inspire type-token ratio ($TT(N)$), where N is the number of tokens, is low (0.025), so we propose that the Inspire vocabulary grows as the vocabulary of unrestricted languages [11]. To prove it, we have investigated the growth of the vocabulary

Table 1: The lexicon growth rate of the Inspire data.

N	96940	293152	364306	535244	614428	716176	809584
$TT(N)$	0.060	0.040	0.033	0.029	0.027	0.026	0.025
$P(N)$	0.027	0.018	0.015	0.013	0.012	0.011	0.010
$S(N)$	0.008	0.005	0.004	0.0037	0.0035	0.0033	0.0031

and its convergence with respect to the sample size. We add information on rare words based on *hapax legomena* ($V(1,N)$) and *dis legomena* ($V(2,N)$) [19]; $V(i,N)$ is the number of types that occur i times in the text with N tokens. We calculate the growth rate of the vocabulary

$$P(N) = \frac{V(1,N)}{N}$$

and Sichel's characteristic

$$S(N) = \frac{V(2,N)}{N}$$

As we can see in Table 1, new words are steadily being added at every stage. More than that, the vocabulary grows approximately at the same rate through the data, and Sichel's characteristic converges as expected [19].

3 Identifying Negotiation Success in Texts

Previous studies on classifying e-negotiations did not consider the language aspect of negotiations. Working with Inspire data, Kersten and Zhang [7] used data mining to classify 1525 negotiations as success or failure. Negotiations were classified using non-language data. Each negotiation was represented by the numbers of offers sent, regularity with which offers were sent, time when the offers were sent, with special attention paid to the time of the last offer, and so on. The average accuracy was 58% for neural networks, 61% for loglinear regression, and 75% for decision trees with the baseline 53%.

Aiming to determine success from the negotiators' language, we are interested in text data that provide substantial information about negotiations. Some negotiations were either abandoned from the start (and failed later), had only official offers and no messages, or had only one message. Obviously, negotiations with one or no message cannot provide useful text information about the negotiation process, so we have deleted them. This eliminated some contradictory examples – when different classifiers treated the same negotiation (with most text characteristics set to 0) as successful or as failed. The amount of noise in the data has been reduced. This clean-up left 1273 negotiations, 761 of them successful.

In this study we work with unedited data, keeping all capitalization versions, spell and grammatical errors made by negotiators. We concatenate the messages that accompany the same negotiation and work with it as one entry. We concentrate on finding language patterns representative of successful (failed) negotiations, that at the same time cannot represent failed (successful) negotiations. For example, *I can accept* is used in successful negotiations almost three times more than in failed negotiations; *you accept my* is used in failed negotiations twice as much as in successful negotiations. This means that we are looking for features frequent in successful (failed) negotiation data and rare or absent in failed (successful) negotiation data. In order to find such features we separate successful from failed negotiations, construct two data sets (successful and failed negotiations, respectively), and investigate the data. For each data we

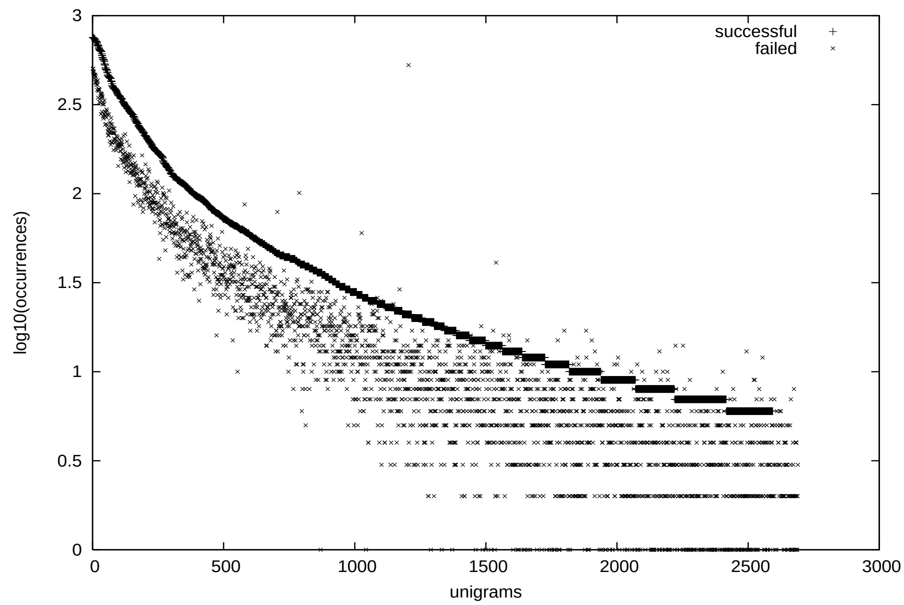

Figure 1: Successful and failed negotiations

build lists of unigrams, bigrams and trigrams. We rank N-grams according to their frequencies starting with the highest frequencies. The rank of a N-gram shows how often it is used in the data relative to other N-grams. By comparing ranks of the same N-gram obtained from successful and failed negotiation data we find which N-grams are frequently present in one of them and rarely present or absent in the other. These N-grams are easily detected when we plot N-grams from successful and failed negotiations; see Fig. 1. The N-grams with a large difference in ranks are depicted as the outliers. Note that the graph for 761 successful negotiations lies, predictably, above the graph for 512 failed negotiations. We have also compared the occurrences of 300 most frequent unigrams in our data when they are used by all negotiators, only by buyers, and only by sellers. We have not found significant differences in the distribution of these unigrams. The similarity of distributions can be seen in Fig. 2.

In Table 2 we list *content words* related to negotiation in general, business and the topic of negotiation. These words are among the 100 most frequent on the list of unigrams from successful negotiations. Their ranks on the list for successful negotiations are rank_s, for failed negotiations – rank_f.

The rank difference rank_s - rank_f is the highest for *agree* (-29), *policy* (-23), *agreement* (-20). It is worth noting that only two of the content words are present on the list of 100 frequent unigrams from failed negotiations and not among the top 100 unigrams from successful negotiations: *find* (rank_s = 134, rank_f = 96), *send* (rank_s = 112, rank_f = 99).

We state our **first hypothesis**: the use of negotiation-related words is relevant to the success of negotiations. To prove the hypothesis we run classification experiments to classify negotiation as successful or failed based on the use of 123 most frequent negotiation-related words. For each negotiation we build a bag of those words [3] by counting how many times each of them is used. We also count the remaining words and add this number to the bag, so each negotiation is represented by a vector of 124 integer-

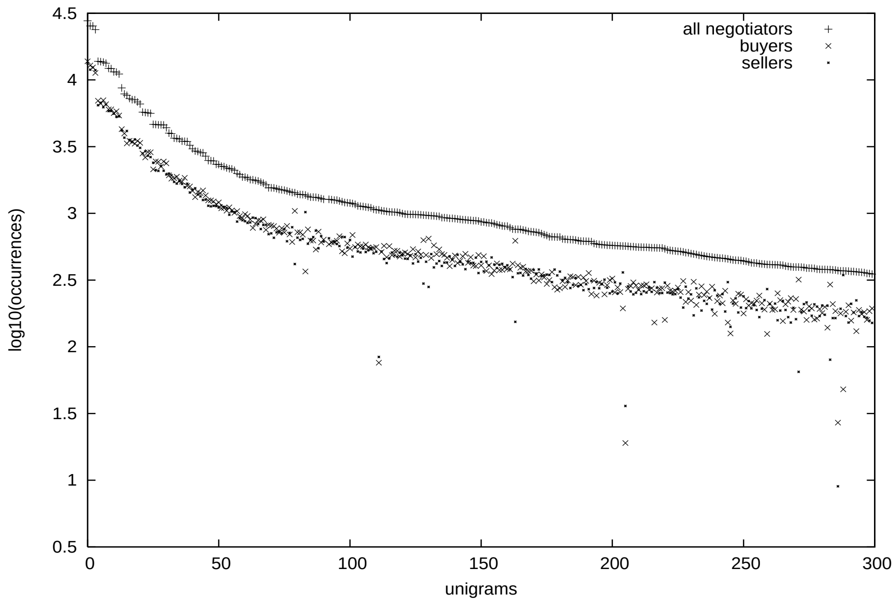

Figure 2: Buyers and sellers

Table 2: Unigrams of the most frequent content words.

unigram	$rank_s$	$rank_f$	unigram	$rank_s$	$rank_f$
offer	8	8	returns	63	69
price	17	20	business	65	65
delivery	26	32	terms	67	59
accept	33	41	agreement	71	91
days	38	37	agree	78	107
think	41	48	Cypress	83	70
payment	42	44	products	84	89
time	43	40	deal	85	87
company	49	46	Itex	86	74
negotiation	53	57	get	89	80
quality	54	52	want	90	94
know	56	47	policy	93	116
make	58	53	give	100	102

Table 3: Classification of negotiations.

Measure	BS	ADT	DLM	DS	DT	IBK	SMO
Precision	60 %	67.4 %	67.2 %	68.2 %	68.1 %	70.3 %	72.9 %
Recall	100 %	84.2 %	90.5 %	83 %	85.2 %	70.4 %	72.3 %

valued features and a binary-valued feature representing the success of negotiation. To classify success or failure of negotiations we have employed several classifiers freely available in the Weka suite [20]: AD Trees (ADT), Decision Stumps (DS), Decision Tables (DT), Instance-based using 20-nearest neighbour (IBK), analog of Support Vector Machine (SMO)as well as the Decision List Machine (DLM) [15]. We have used 10-fold cross-validation to perform experiments. Our experiments have resulted in 66-69% overall accuracy of classification with the baseline(BS) 60%. Precision and recall results for classifying negotiations are reported in Table 3.

Simons [17] found that language patterns of the first part of negotiation efficiently predict the negotiation outcome. In our data both company names, *Cypress* and *Itex*, have higher ranks in failed than in successful negotiations. Recall from Section 2 that a buyer represents *Cypress* and a seller represents *Itex*[1] We note that the negotiators mostly use the company name when they introduce themselves to the partner. The difference in the company name ranks points to the fact that the language of successful negotiations differs from the language of failed negotiations from the very beginning of the process. We state our **second hypothesis**: the starting phase of successful negotiations is different from the starting phase of failed negotiations.

To prove the second hypothesis we first calculate that in successful negotiations *Cypress* and *Itex* together account for 3/4 of the data they account for in failed negotiations. We also compare ranks of highly frequent bigrams and trigrams containing company names that appear among 500 most frequent bigrams and 700 most frequent trigrams. *Cypress* appears in successful negotiations in one bigram (*Cypress Cycles*, $rank_s = 96$), two trigrams (*Cypress Cycles Dear*, $rank_s = 467$; *of Cypress Cycles*, $rank_s = 532$). In failed negotiations *Cypress* appears in one bigram (*Cypress Cycles*, $rank_f = 55$), five trigrams (*represent Cypress Cycles*, $rank_f = 359$; *Cypress Cycles is*, $rank_f = 391$; *I represent Cypress*, $rank_f = 420$; *of Cypress Cycles*, $rank_f = 474$; *Cypress Cycles Dear*, $rank_f = 625$). *Itex* appears in successful negotiations in one bigram (*Itex Manufacturing*, $rank_s = 285$), and in no trigrams. In failed negotiations *Itex* appears in one bigram (*Itex Manufacturing*, $rank_f = 295$), and in no trigrams.

The comparison of the N-gram ranks leads us to conclude that in both successful and failed negotiations *Cypress* is used more frequently than *Itex*. *Cypress* is noticeably more frequently used in failed than in successful negotiations. The use of *Itex* in successful negotiations is more frequent than in failed ones.

We pay attention to the fact that the trigram *Cypress Cycles Dear* is the first most frequent in successful negotiations and fifth in failed negotiations. The **third hypothesis** that naturally arises from this observation is that politeness is a characteristic feature of successful negotiations. To prove this hypothesis we give the unigram ranks for such indicators of polite speech as *Thank* ($rank_s = 69$, $rank_f = 83$), *Thanks* ($rank_s = 108$, $rank_f = 109$), *thank* ($rank_s = 250$, $rank_f = 315$), *thanks* ($rank_s = 420$, $rank_f = 379$). We deliberately preserve case-sensitivity, to enable further studies of negotiators' attitude towards the negotiation process and negotiation partners. In the present study, however, we calculate the combined percentage of the orthographic variations of the words *thank*, *thanks*. Remarkably, it is 2.5 times more in successful than in failed negotiations!

In Table 4 we report the ranks of most frequent trigrams with the politeness indicator words. Pre-

[1] The role-related issues are the subject of studies of the forthcoming paper.

Table 4: The Most Frequent Trigrams with *Thank*, *Thanks*, *thank* and *thanks*.

word	N	trigram	rank_s	trigram	rank_f
Thank	1	Thank you for	1	Thank you for	7
	2	Thank you very	103	Thank you very	116
Thanks	1	Thanks for your	29	Thanks for your	27
	2	Thanks for the	609		
thank	1	thank you for	42	thank you for	85
	2	to thank you	469		
thanks	1	thanks for your	491	thanks for your	659

Table 5: Unigrams of the most frequent modal verbs.

unigram	rank_s	rank_f	unigram	rank_s	rank_f
have	15	15	could	103	97
can	23	24	need	104	100
will	24	22	should	152	154
would	37	34	must	184	177

dictably, the negation is presented more often in the failed than in the successful negotiations: *not* ($\text{rank}_s = 30$, $\text{rank}_f = 26$).

We have also compared the use of verbs *have*, *can*, *will*, *would*, *need*, *could*. Verbs found in the top 100 frequent unigrams in both successful and failed negotiations are *have*, *can*, *will*, *would*. The verbs *could, need* are found in the 100 most frequent words from failed negotiations and absent in the top 100 unigrams from successful negotiations. Hence the negotiators who fail to reach an agreement use the modal verbs and the verb *have* marginally more frequently than the negotiators who have succeeded in reaching a compromise. To support this claim in Table 5, we add the ranks for two modal verbs: *should*, *must*.

We employ the bigrams to find more language differences between successful and failed negotiations. We locate the bigrams, containing the above listed content words, in the lists of 100 most frequent bigrams from successful negotiations, and present them in Table 6. The highest difference in ranks (41) is obtained on *Cypress Cycles*. "Cypress Cycles" is the name of the buying company, used mainly during the introductory phase. This again supports our second hypothesis that the introductory phase in successful negotiations is different than in failed negotiations.

We compare the ranks of the most frequent bigram (both in successful and failed negotiations) with negation – *is not* ($\text{rank}_s = 95$, $\text{rank}_f = 65$), and the most frequent bigram (again both in successful and failed negotiations) with an agreement indicator – *agree with* ($\text{rank}_s = 157$, $\text{rank}_f = 244$). As expected, the "negative" bigram is more frequent in failed negotiations, and the "positive" bigram in successful negotiations.

4 Finding Language Patterns

Our next step is to find trigrams highly indicative of success or failure of negotiations. As we said before, these trigrams have a large difference between their ranks in successful and failed negotiations. We are looking for trigrams that show the negotiators' goal (win by any means, reach a compromise, do away

Table 6: Bigrams of the most frequent content words.

bigram	$rank_s$	$rank_f$	bigram	$rank_s$	$rank_f$
your offer	2	2	to accept	57	70
I hope	4	7	your company	70	59
the price	5	10	an agreement	67	97
offer I	24	30	offer is	71	78
like to	29	17	the delivery	74	72
offer and	32	44	the payment	76	96
new offer	37	65	price is	82	122
my offer	39	31	upon delivery	83	113
this offer	41	60	full price	84	110
last offer	45	48	return policy	86	132
accept your	48	69	price and	88	98
hope you	51	41	delivery time	90	90
hope that	52	62	the offer	91	83
to make	56	37	Cypress Cycles	96	55

with the assignment), their attitude to partners (friendliness, aggressiveness, indifference), and behaviour in the negotiation process (flexibility, stubbornness). The same trigrams must be noticeably present in either successful or failed negotiations. Hence two major elements affect the N-gram selection:

1. words contained in an N-gram,
2. its rank.

A procedure of building the lists of language patterns
for successful and failed negotiations

Input: text data of all negotiations, text data of successful negotiations, text data of failed negotiations.

1. Build the list L of unigrams for all negotiations
2. Build the lists of N-grams ($N = 1, 2, 3$) for successful negotiations (SN)
3. Build the lists of N-grams ($N = 1, 2, 3$) for failed negotiations (FN)
4. In L find unigrams of nouns and non-modal verbs among k most frequent unigrams from the lists of negotiation, business and topic of negotiation zones. Build the list W of such nouns and verbs.
5. For each $w \in W$:
   - Find its rank r_s^1 in the list of the unigrams of SN.
   - Find its rank r_f^1 in the list of the unigrams of FN.
   - Find $d_w^1 = r_s^1 - r_f^1$.
6. Delete from W all w such that $d_w^1 < d$.

7. For each $w \in W$:
   - Find its bigrams among m most frequent bigrams on the list of bigrams of SN.
   - Find its bigrams among m most frequent bigrams on the list of bigrams of FN.
   - – Find the rank r_s^2 of the i_{th} bigram on the list of the bigrams of SN.
     – Find the rank r_f^2 of the i_{th} bigram on the list of the bigrams of FN.

     – Find $d_i^2 = r_s^2 - r_f^2$.
   - Calculate $d_w^2 = \sum_{i=1}^{m} d_i^2$
8. Delete from W all w such that $d_w^2 < d$.
9. Find most frequent trigrams containing unigrams from W: repeat steps 7-8 for trigrams instead of bigrams.
10. Build the list L_R of trigrams, containing $w \in W$, with their ranks.

Output: L_R.

In our procedure we use adjustable parameters: the distance d and cut-off points k, m. In order not to overload the procedure with parameters, we do not use weights to tune distances between N-grams, though it seems a natural thing to do. We have tested the procedure with $d = min(100, 2\dot{r}ank_s)$, $k = 100$, $m = 700$. Among different techniques that can be employed to find cut-off parameters k, m we have chosen the one that eliminates low-frequency N-grams and keeps representative N-grams in negotiation data. Needless to say, the choice of distance depends on the cut-off points. For the cut-off points mentioned, the distance we used ensures that the difference in ranks provides difference in N-gram frequencies.

For first bigrams the distance $rank_s$ - $rank_f$ between ranks is the highest in *return policy* (87 - 133 = -46), *Cypress Cycles* (97 - 56 = 41), *an agreement* (68 - 98 = -30), *full price*(84 - 110 = -26), *accept your* (48 - 69 = -21), *the payment* (76 - 96 = -20). Note that for each word these bigrams are the first ones in both lists.

We want to eliminate bigrams that do not help distinguish successful from failed negotiations. We calculate the distances $rank_s - rank_f$ for the second frequent bigrams with a content word. For example in both lists the second bigram with the word *policy* is *returns policy* (425 - 560 = -135). The ranks of the second bigram with *Cypress* fall below the cut-off point (*of Cypress*, $rank_s$ = 715, *with Cypress*, $rank_f$ = 792), and second bigrams with the words *price* and *delivery* are closely ranked in both lists.

After calculating distances we concentrate our search on the trigrams originating from *policy, agreement, accept, payment, not, agree*. These words we consider to be seeds of the trigrams that distinguish successful from failed negotiations. We seek three most frequent trigrams originated from each of the seeds. After calculating distances between trigrams of the same seed, we delete trigrams originating from *agreement* from the lists of distinguishing trigrams. Based on the bigram analysis, we also delete the trigrams containing the word *delivery*. The resulting list is presented in Table 7.

We notice that in the trigrams from the failed negotiations there is a trace of aggressive behaviour (**you will accept, you will agree, you are not**), which is absent from the corresponding trigrams in the successful negotiations (**you can accept, agree with your, it is not**). Tracing the trigrams with "you", we found that in successful negotiations they correspond to politeness, in failed negotiations – to aggressiveness.

Table 7: The List of Representative Trigrams

word	N	trigram	rank$_s$	trigram	rank$_f$
policy	1	the return policy	65	the return policy	82
	2	the returns policy	494	return policy I	594
	3	return policy is	538	price return policy	820
accept	1	accept your offer	14	accept your offer	26
	2	to accept your	55	to accept your	103
	3	you can accept	90	you will accept	132
agree	1	agree with you	361	agree with the	509
	2	agree with your	395	you will agree	533
	3	I agree with	426	agree with you	565
payment	1	payment and returns	225	terms of payment	223

Our next goal is to investigate the different ways in which modal verbs and the verb *have* appear in successful and failed negotiations. We look for patterns most frequent in their classes. Noticeably, the verbs "have" and "will" appear more frequently in failed than in successful negotiations, while the verb "can" is more frequent in successful than in failed negotiations. See Table 8 for results.

To find patterns indicative of success or failure, we look for trigrams frequent in one class and rare in another class of negotiations. We report the results in Tables 9, 10 and 11. In Table 9 a negative distance means a trigram more frequent in successful negotiations, a positive distance – a trigram more frequent in failed negotiations.

Consider the distances of trigrams with the pronoun "I": we posit that the negotiators' assertiveness is higher in successful than in failed negotiations. The same conclusion about assertiveness can be drawn from the trigrams presented in Tables 10 and 11. As expected, trigrams that suggest uncertainty and rejection are indicative of failed negotiations.

5 Conclusions and Future Work

In this paper we have presented a new procedure of identifying language patterns indicative of the outcome of negotiations. Applying Natural Language Processing and Machine Learning technique, we have proved that text messages exchanged by the users of a negotiation support system are relevant to the outcome of the e-negotiations. We have related language and the success of negotiations by arguing in favour of three hypotheses.

We have shown that e-negotiations can be classified correctly if they are represented by bags of words built from negotiation-related words. Our classification results, based only on language, are comparable with the classification results achieved on the non-language data.

We have stated and shown how the language differs in successful and failed negotiations. We have demonstrated that the language of the initial phase of successful e-negotiations is different than in failed e-negotiations. We have shown that politeness is an essential part of successful negotiations.

In this paper we did not emphasize differences between the language patterns of buyers and sellers. The role-dependent patters are very interesting and promising direction of studies and are left for the future work.

In the future we plan to incorporate Machine Learning methods in the procedure of finding language

Table 8: The most frequent trigrams with verbs have, can, will.

word	N	trigram	rank$_s$	trigram	rank$_f$
have	1	I have to	49	I have to	36
	2	that we have	55	we have to	43
	3	we have to	66	that you have	75
	4	that you have	92	that we have	92
	5	to have a	123	will have to	162
can	1	that we can	16	that we can	16
	2	hope we can	62	hope we can	74
	3	hope you can	73	hope you can	99
	4	you can accept	89	if you can	117
	5	that you can	118	that you can	147
will	1	that you will	36	that you will	22
	2	hope you will	53	hope you will	30
	3	I will be	69	you will find	44
	4	that we will	83	I will be	52
	5	will be able	85	we will be	93

Table 9: Trigrams with largest differences in ranks

word	trigram	d^3	trigram	d^3
have	we will have	-213	you have any	-163
	if you have	-119	offer I have	96
can	we can come	166	that I can	-106
	can come to	95	you can accept	-94
cannot	I cannot accept	-97	we can not	127
	can not accept	51		
will	and I will	125	will not be	67
	you will be	-52		
would	we would like	52	I would be	-28

Table 10: Trigrams presented in 450 top trigrams of successful negotiations and absent in 450 top trigrams of failed negotiations

word	trigram	rank$_s$	trigram	rank$_s$
have	that I have	284	I have made	329
	and I have	355	we can have	360
	I have a	390		
can	I can accept	172	as you can	350
will	hope we will	187	I will accept	287
	offer will be	381		
would	we would be	343		

Table 11: Trigrams presented in 450 top trigrams of failed and absent in 450 top trigrams of successful negotiations

word	trigram	rank_f	trigram	rank_f
have	have received your	194	I have not	288
	I have received	313	I have already	383
can	if we can	234	we can do	256
	so we can	315	we can work	338
	think we can	362		
cannot	can't accept your	201	we cannot accept	320
will	will find it	250	will be a	341
	that I will	386		
would	you would like	359		

patterns. How Machine Learning methods are implemented poses questions about different mappings of negotiations to bags of words. This is closely interrelated with the question how the use of non-negotiation related words affects the negotiation outcome.

In this paper we have avoided an investigation of the dependence of text noise and the e-negotiation process. This is also left for the future work. Preliminary studies show that there is correlation between non-negotiation related text data, noise level and the outcome of negotiations.

Acknowledgment

This work is partially supported by the Social Sciences and Humanities Research Council of Canada and by the Natural Sciences and Engineering Research Council of Canada.

References

[1] S. Jekat, A. Klein, E. Maier, I. Maleck, M. Mast, J. Quantz. "Dialogue Acts in VERBMOBIL", *VM-Report 65*, 1995.

[2] K. Jokinen, T. Hurtig, K. Hynna, K. Kanto, M. Kaipainen, A. Kerminen. "Self-Organizing Dialogue Management", *Proc 2nd Workshop on Neural Networks and Natural Language Processing (NLPRS)*, Tokyo, Japan, 2001.

[3] D. Jurafsky and J. H.Martin. *Speech and Language Processing*. Prentice Hall (2000).

[4] G. E. Kersten. "The Science and Engineering of E-negotiation: An Introduction". InterNeg Report 02/03, 2003. interneg.org/interneg/research/papers/

[5] G.E. Kersten. Modeling Distribution and Integrative Negotiations. Review and Revised Characterization. *Group Decision and Negotiation*, **10**(6), 493-514, 2001.

[6] G. E. Kersten and S. J. Noronha. "WWW-based Negotiation Support: Design, Implementation, and Use". *Decision Support Systems*, **25**, 135-154, 1999.

[7] G. E. Kersten and G. Zhang. "Mining Inspire Data for the Determinants of Successful Internet Negotiations", INR04/01, 2001. To appear in *European Journal of Operational Research*, 2003.

[8] C. D.Manning, H. Schutze. *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.

[9] M. Mast, H. Niemann, E. Noth, E. G. Schukat-Talamazzini. "Automatic Classification of Dialog Acts with Semantic Classification Trees and Polygrams". *Learning for Natural Language Processing*, Springer, **1040**, 217-229, 1996.

[10] M. Morris, J. Nadler, T. Kurtzberg, L. Thompson "Schmooze or Lose: Social Friction and Lubrication in E-mail Negotiations", **6**, Group Dynamics, 89-100, 2002.

[11] T. McEnery, A. Wilson. *Corpus Linguistics.* Edinburg University Press, 2001.

[12] M. P. Oakes. *Statistics for Corpus Linguistics.* Edinburg University Press, 1998.

[13] C. Quix, M. Schoop , M. Jeusfeld. "Business Data Management for Business-to-Business Electronic Commerce". *SIGMOD Record*, **31**(1), 49-54, 2002.

[14] M. Schoop. "A Language-Action Approach to Electronic Negotiations". *Proc 8th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2003)*, 143-160, 2003.

[15] M. Sokolova, M. Marchand, N. Japkowicz, J. Shawe-Taylor "The Decision List Machine", *Advances in Neural Information Processing Systems*, **15**, 921-928, The MIT Press, 2003.

[16] M.Sokolova, S. Szpakowicz, V. Nastase. "Automatically Building a Lexicon from Raw Noisy Data in a Closed Domain". Submitted to *Information Intelligent Systems-2004*.

[17] T. Simons. "Speech Patterns and the Concept of Utility in Cognitive Maps: the Case of Integrative Bargaining". *Academy of Management Journal*, **38**(1), 139-156, 1993.

[18] L. Thompson, J. Nadler. "Negotiating Via Information Technology: Theory and Application", *Journal of Social Issues*,**58**(1), 109-124, 2002.

[19] F. J. Tweedie, R. H. Baayen. "How Variable May a Constant be? Measures of Lexical Richness in Perspective". *Computers and the Humanities*, **32**, 323-352, 1998.

[20] I. Witten, E. Frank. *Data Mining*, Morgan Kaufmann, 2000. http://www.cs.waikato.ac.nz/ml/weka/