Table 3 A argumentation tag annotation of a salesperson's utterance

`<arg alt=A><fra alt=A,polarity=POS,pref=NO>`(Camera A is) able to achieve performance of comparable single-lens cameras and can fit in your pocket `</fra>`, this is a point.`</arg>`

$$\texttt{<arg alt=}a_j\texttt{>} \ ... \ \texttt{</arg>} \tag{5}$$

We also define annotated framing tags F as follows:

$$F = \{f_1, f_2, ..., f_I\} \tag{6}$$
$$f_i = \langle a_i, p_i, r_i \rangle \tag{7}$$
$$a_i \in ALT \tag{8}$$
$$p_i \in \{\text{POS}, \text{NEG}\} \tag{9}$$
$$r_i \in \{\text{YES}, \text{NO}\} \tag{10}$$

where a_i represents the target alternative, p_i takes value NEG if the framing is negative, and POS if the framing is positive, and r_i represents whether the arguments contain a reference to the persuadees preferred determinant, taking the value TRUE if contained, and FALSE is not contained. The user's preferred determinant is annotated on the basis of the results of questionnaire. In the annotated corpus, f_i is described by the following format:

$$\texttt{<fra alt=}a_i\texttt{,polarity=}p_i\texttt{,pref=}r_i\texttt{>} \ ... \ \texttt{</fra>} \tag{11}$$

Table 3 shows an example of annotation of positive framing (p=POS) about the performance of Camera A (a=A). In this example, the customer answered that his preference is the price of camera, and this utterance does not contain any description of price. Thus, r=NO is annotated. Finally, we annotate `<arg alt=A>` around the entire utterance because at least one `fra` tag exists.

3.2.2 Reliability of Annotation

To evaluate the reliability of the annotation, we randomly selected 10% of the collected data and evaluated the data for inter-annotator agreement. The GPF and argumentation tags were evaluated on the basis of the agreement between two annotators. The description section and the variables of the `fra` tag were evaluated by a second annotator regarding whether the annotation result of the primary annotator was acceptable or not. The acceptability rate is calculated as the percentage of tags judged as appropriate by the second annotator out of the tags annotated by the primary annotator.

Initially, the agreement of the 18 annotated GPF tags was only 30%. As this is too low to achieve reliable results in our analysis, we merged tags with low agreement, resulting in a total of 6 tags and an agreement of 76% (see Table 4). This agreement is comparable to other research in a different task [16]. We use these merged GPF

Table 4 Result of the merging GPF tags

GPF tag	PROPOSITIONALQ, SETQ, INFORM, ANSWER, COMMISSIVE, DIRECTIVE

tags in the analysis of later sections. The agreement of the argumentation tag was 94%, and, the acceptability rate of the description section and target candidate of the framing tag was 94%, polarity was 100%, and of preference information was 82%.

4 Success Measures and Dialogue Factors

Given the corpus described in the previous section, we would now like to elucidate the factors that contribute to persuasive power and user satisfaction.

4.1 Success Measures for Persuasive Dialogue

First we define our measures for success of persuasive dialogue. As the dialogue consists of two interlocutors, we define a successful dialogue as a dialogue where both participants achieve their goal. As with dialogue systems, simply using satisfaction as measure of dialogue success for the persuadee seems appropriate [17]. However, as far as we are aware there is no widely shared evaluation criterion in the relatively young field of persuasive technology. Thus we propose two measures for the success for the persuader: 1) Whether the persuadee finally chooses the persuasive target at the end of the dialogue, and 2) the amount the persuadees intention changed about the persuasive target as the result of the dialogue.

We measured these values by conducting a questionnaire of the persuadees to measure satisfaction, intention change about the persuasive target, and success of persuasion, as described below:

Satisfaction (*Sat*): The persuadee's subjective satisfaction with the persuader defined as a 5 level score of customer satisfaction (1: Not satisfied, 3: Neutral, 5: Satisfied).

Intention change (ΔIn): The amount the persuadees intention to buy the persuasive target changed as a result of the dialogue. We conducted a questionnaire about intention to buy persuasive target (1: Don't want to buy, 3: Neutral, 5: Want to buy) before (In_{before}) and after (In_{after}) the dialogue. ΔIn is measured as follows:

$$\Delta In = In_{after} - In_{before} \tag{12}$$

Persuasive success (*Suc*): *Suc* takes the value 1 when the customer decides to purchase the persuasive target at the end of dialogue, and 0 otherwise.

4.2 Dialogue Factors

In this section, we describe several measurable characteristics of the dialogue that may contribute to persuasive power and user satisfaction. These include factors regarding negative/positive framing, original preference of the persuadee, and dialogue acts.

4.2.1 Factors Regarding Negative/Positive Framing

Two dialogue factors to measure negative-positive framing are defined as follows:

Negative framing ratio for non-target ($R_{\mathrm{NEG},a\neq t}$): The ratio of utterances stating negative facts about alternatives other than the persuasive target, where we define t as the persuasive target:

$$R_{\mathrm{NEG},a\neq t} = \frac{\sum_{k=1}^{K} \delta(\exists_{f\in u_k.F}(f.a \neq t \wedge f.p = \mathrm{NEG}))}{\mathrm{K}}, \tag{13}$$

where δ is Kronecker's delta, 1 when the condition is true, and 0 otherwise.

Positive framing ratio for target ($R_{\mathrm{POS},a=t}$): Likewise the ratio of utterances by the persuader positively framing the persuasive target:

$$R_{\mathrm{POS},a=t} = \frac{\sum_{k=1}^{K} \delta(\exists_{f\in u_k,F}(f.a = t \wedge f.p = \mathrm{POS}))}{\mathrm{K}}. \tag{14}$$

4.2.2 Factors Regarding the Persuadees Original Preference

We also define 3 kinds of factors to measure the persuadees attitude change.

Conveyed preferred determinant (CPD_a): Whether the persuadee has been told by the persuader that alternative a satisfies the determinant that the persuadee has mentioned as important in the pre-dialogue questionnaire

$$CPD_a = \delta(\exists_f f =< a, \mathrm{POS}, \mathrm{YES} >). \tag{15}$$

Prior candidate evaluation (PCE_a): The persuadees evaluation of alternative a at the beginning of dialogue. In this paper, we calculated one feature for each alternative that is 1 if that alternative is selected by the persuadee as preferred before the dialogue and 0 otherwise.

Prior persuasive target evaluation ($PPTA$): The persuadees evaluation of the persuasive target at the beginning of the dialogue as measured by questionnaire.

4.2.3 Other Factors

In addition to the above factors, we defined factors based frequency of traditional dialogue acts and argumentation, and total time.

Number of argumentation events (I): The total number of occurrences of argumentation tags during the dialogue I.

Prodeecings of 5th International Woskhop on Spoken Dialog Systems
Napa, January 17-20, 2014

Table 5 Distribution of general purpose function (GPF) and argumentation tags

	GPF						Argument		
	PropQ	SetQ	Commisive	Directive	Answer	Inform	Tar	NonTar	Both
Salesperson	14%	4%	6%	8%	16%	45%	25%	3%	3%
Customer	21%	2%	9%	5%	17%	37%	-	-	-

Frequency of general purpose function ($R_{r,g}$): The ratio of each GPF tag for each role in the dialogue

$$R_{r,g} = \frac{\sum_i^I \delta(u_{i=1} = \langle r, g, \bullet, \bullet \rangle)}{\sum_i^I \delta(u_{i=1} = \langle r, \bullet, \bullet, \bullet \rangle)}. \tag{16}$$

Total time (TT): Total dialogue time in seconds.

5 Analysis

In this section, we present a manual analysis of the dialogue acts included in the corpus, and a linear regression analysis of the factors that contribute to persuasion.

5.1 Analysis of Dialog Acts

First, in order to perform a general analysis of the main dialogue acts comprising persuasive dialogue, we show the proportion of argumentation tags of all utterances of the salesperson and the GPF distribution for both the customer and salesperson in Table 5. From the result, we can see information presentation (Answer, Inform) tags cover more than half of both of the customer and salesperson utterances. In addition, when considering information seeking tags (PropQ, SetQ), the percentage reaches about 80%.

31% of all dialogue acts of the salesperson are arguments. This indicates that the argumentation tag proposed in Section 3 is highly relevant in this situation. A more detailed breakdown is that 25% of arguments target only the persuasive target, 3% of arguments target only an alternative other than the persuasive target, and 3% of arguments target both the persuasive target and a non-persuasive target. This indicates that, in persuasive dialogue, the persuader rarely suggests arguments for selecting alternatives other than persuasive target, but does occasionally mention other options.

Table 6 shows mean persuadee satisfaction categorized by initial and final choice of alternative. The results seem to indicate that it is possible to achieve satisfaction and persuasion simultaneously when the customer has initially chosen the persuasive target or doesn't have an initial choice, but it is harder when the customer has initially chosen an alternative other than the persuasive target. However, the data is still somewhat small to make conclusions about this fact.

Prodeecings of 5th International Woskhop on Spoken Dialog Systems
Napa, January 17-20, 2014

Table 6 Average satisfaction (and number of dialogues) for each initial and final choice.

		Final Choice		
		PT	Not PT	None
Initial Choice	PT	4.0(3)	-	5.0(5)
	Not PT	2.0(2)	2.0(1)	4.4(7)
	None	4.0(3)	2.0(2)	3.4(7)

Table 7 Linear regression for satisfaction and persuadees intension change, and logistic regression for success of persuasion with selected factors. All factors are normalized.

	$w_0 + w_1 x_1 + ... w_n x_n$						R^2
Sat	+3.56	*Bias*	+.501	$R_{\text{SALES,PROPQ}}$	-.509	$R_{\text{SALES,COMMISIVE}}$	.396
ΔIn	+.920	*Bias*	-.475	$R_{\text{NEG},a \neq t}$	+.625	I	.640
	-.303	CPD_E	+.429	$PPTA$	+.295	PCE_C	
	+.422	$R_{\text{CUST,ANSWER}}$	+.464	$R_{\text{CUST,INFO-PROV}}$	+.276	$R_{\text{CUST,COMMISIVE}}$	
	-.368	TT					
	$w_0 + w_1 x_1 + ... w_n x_n$						Accuracy
Suc	-4.349	*Bias*	+2.00	CPD_B	-8.14	PCE_B	80%
	-2.12	TT					

5.2 Regression Analysis of Factors in Persuasion

To analyze the relationship between the success measures in Section 4.1 and factors in Section 4.2, we performed a regression analysis to discover the important factors and measure accuracy of the prediction model. Factor selection is performed using step-wise multinomial linear regression [18]. We repeatedly perform multinomial regression and exclude predictors that do not sufficiently contribute to the model until we get a model for which all of the predictors are significant. In this research, we excluded any predictor with a p-value above .25 at each iteration, and the final model is comprised of predictors that are statistically significant ($p < .05$). Prediction accuracy of the selected factors is evaluated through leave-one-out cross validation after the selection.

Table 7 shows the results. First focusing on the factors for satisfaction, we can see that predictors account for 39% of the variance of satisfaction. Focusing on the variables selected as useful in the linear regression results, we can see that both of the two features come from the salesperson's GPF tags. The weight of $R_{\text{SALES,PROPQ}}$ is high, which indicates that by asking many questions, the salesperson can make the customer feel more satisfied with the conversation. The reason why the weight of $R_{\text{SALES,COMMISIVE}}$ is assigned a large negative value is that $R_{\text{SALES,COMMISIVE}}$ represents the degree of failure in answering the customer's questions. For example, most of the utterances such as "Sorry, I don't know. I'll take a look" are annotated COMMISIVE. This result is interesting, as it shows that customer satisfaction is largely dependent on the salesperson, a fact that may guide our implementation.

Next, focusing on the weight of factors in the linear regression results for opinion change, factors derived from argumentation tags account for 46% of total weight, making the largest contribution to prediction. The highest weight is I, indicating that more argumentation for the persuasive target results in a larger change in the opinion of the persuadee. On the other hand, *PPTA* is assigned a large negative weight, indicating the persuader does not change the opinion of a persuadee who already wanted to select the persuasive target a priori, a natural result as the persuader will not want to change an already favorable result. The weight of factors derived from the GPF tag account for 33% of the total weight. Especially, the ratio of information-exchange ($R_{\text{CUST,ANSWER}}$, $R_{\text{CUST,INFO-PROV}}$) assumes a high weight, indicating that making the customer speak more contributes to opinion change.

Finally, looking at the result for logistic regression over persuasive success, we can see that 80% of the data are correctly predicted, compared to a chance rate of 68% when predicting only failure of persuasion. Focusing on the weights of the variables in the logistic regression result, the weight of PCE_B is relatively high, indicating that if customers select camera B pre-dialogue, the persuasion becomes more difficult. CPD_B is the only variable with positive weight, indicating that informing the persuadee about alternatives other than the persuasive target that match the persuadee's preference increases the persuasive power for the persuasive target. We hypothesize the reason why only camera B appeared in predictors is that camera B was chosen many times compared to other alternatives, and appeared as the alternative for comparison to the persuasive target in many dialogues.

Combining all these results together, we can see that the persuader is required to use a sophisticated dialogue strategy, as different factors contribute to the achievement of successful persuasion and persuadee satisfaction. However in Table 7, we can also see that no predictor influences both successful persuasion and persuadee satisfaction. Therefore, the persuader could potentially perform dialogue to achieve both goals simultaneously. For example, the persuader would perform a large amount of argumentation to achieve persuasion, and ask many questions to increase user satisfaction. However, as observed by the negative weight for TT, intention change of the persuadee also tends to decrease as time passes. Thus, the persuader must achieve both goals in a short time, considering interaction efficiently and accurately predicting the persuadees interest in each of the alternatives.

6 Conclusion

In this paper, we analyzed persuasive dialogue between humans, focusing on the factors that contribute to persuasion and satisfaction. In order to do so, we collected a corpus of dialogues between salespeople and customers, and defined an argumentation tag scheme and dialogue factors for predicting dialogue goals.

The experimental results indicate that the main dialog acts that compose the dialogue are information exchange and argumentation. A regression analysis demonstrated that argumentation contributes effectively to the achievement of persuasion, and factors derived from GPF were effective for predicting satisfaction.

Our next step in this research is to incorporate these observations into the persuasive dialogue framework of [9]. In addition, this experiment result is still limited in the corpus we collected. We will investigate the flexibility of the proposed tag scheme and persuasive factors on other persuasion tasks.

Acknowledgements Part of this work was supported by the collaborative research with I2R and ATR-Trek.

References

1. Wayne Ward, "The CMU air travel information service: Understanding spontaneous speech," *Proceedings of the Workshop on Speech and Natural Language*, 1990.
2. Teruhisa Misu, Komei Sugiura, Kiyonori Ohtake, Chiori Hori, Hideki Kashioka, Hisashi Kawai, and Satoshi Nakamura, "Modeling spoken decision support dialogue and optimization of its dialogue strategy," *ACM Transactions on Speech and Language Processing*, 2011.
3. Joseph Weizenbaum, "ELIZA – a computer program for the study of natural langage communication between man and machine," *Communications of the Association for Computing Machinery*, 1966.
4. B.J.Fogg, *Persuasive Technology*, Morgan Kaufman, 2003.
5. Anton Nijholt, Ronald C Arkin, Sebastien Brault, Richard Kulpa, Frank Multon, Benet Bideau, David Traum, Hayley Hung, Eugene Santos Jr, Dequing Li, Fei Yu, Lina Zhou, and Donsong Zhang, "Computational deception and noncooperation," *IEEE Intelligent Systems*, 2012.
6. Stephen Purpura, Victoria Schwanda, Kaiton Williams, William Stubler, and Phoebe Sengers, "Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
7. Marco Guerini, Oliviero Stock, and Massimo Zancanaro, "Persuasion model for intelligent interfaces," *Proceedings of the IJCAI Workshop on Computational Models of Natural Argument*, 2003.
8. Kallirroi Georgila and David Traum, "Reinforcement learning of argumentation dialogue policies in negotiation," *Proceedings of INTERSPEEECH*, 2011.
9. Takuya Hiraoka, Yuki Yamauchi, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura, "Dialogue management for leading the conversation in persuasive dialogue systems," *Proceedings of ASRU*, 2013.
10. Kallirroi Georgila, Ron Arstein, Angela Nazarian, Michael Rushforth, David Traum, and Katia Sycara, "An annotation scheme for cross-cultural argumentation and persuasion dialogues," *Proceedings of the SIGDIAL*, 2011.
11. Sandra L. Schneider Levin Irwin and Gary J. Gaeth, "All frames are not created equal: A typology and critical analysis of framing effects," *Organizational behavior and human decision processes 76.2*, 2013.
12. Hien Nguyen, Judith Masthoff, and Pete Edwards, "Persuasive effects of embodied conversational agent teams," *Proceedings of Human-Computer Interaction*, 2007.
13. ISO 24617-2, *Language resource management-Semantic annotation frame work (SemAF), Part2: Dialogue acts. ISO*, 2010.
14. Martin Fishbein, "An investigation of the relationship between beliefs about an object and the attitude toward that object," *Human relations*, 1963.
15. Irene Mazzotta and Fiorella de Rosis, "Artifices for persuading to improve eating habits," *AAAI Spring Symposium: Argumentation for Consumers of Healthcare*, 2006.
16. Kiyonori Ohtake, Teruhisa Misu, Chiori Hori, Hideki Kashioka, and Satoshi Nakamura, "Annotating dialogue acts to construct dialogue systems for consulting," *Proceedings of the 7th Workshop on Asian Language Resources. Association for Computational Linguistics*, 2009.
17. Marilyn A Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella, "Paradise: A framework for evaluating spoken dialogue agents," *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, 1997.
18. Allison Terrell and Mutlu Bilge, "A regression-based approach to modeling addressee backchannels," *Proceedings of the 13th Annual Meeting of SIGDIAL*, 2012.