A CD-ROM Retrieval System with Multiple Dialogue Agents

Keiichi Sakai, Tsuyoshi Yagisawa, and Minoru Fujita
Canon Inc., Media Technology Laboratory
890-12 Kashimada, Saiwai-ku, Kawasaki, 211, Japan
{keiichi,yag428,minoru}@cis.canon.co.jp

Abstract

In this paper, we proposed a new dialogue system with multiple dialogue agents. In our new system, three types of agents: a) domain agents, b) strategy agents, and c) context agents were realized. They give the following advantages to the user:

- the domain agents make the user aware of the boundary between the domains.
- the strategy agents make the user aware of the difference between the strategies.
- the context agents help the user to deal with multiple goals.

We expect that the complex behaviors of the system will become more visible to the user in different situations. The experimental results show that the user can retrieve effectively and obtain the expected goals easily by using these multiple agents.

1 Introduction

Recently, research into 'multi-agent system' is increasing. The multi-agent system is now one of the promising solutions to achieve a complicated system (Maes, 91; Nishida and Takada, 93; Nagao and Takeuchi, 94).

The multi-agent system which simulates cooperation between 'human-agents' is realized by an integration of simplified autonomous functions. And as a result it achieves a complicated system in total. It also has a latent potential to make a very flexible system.

Thus, we believe that if we introduce the concept of the multi-agent system into a dialogue system, we are able to construct a more sophisticated system which is able to treat various linguistic phenomena and to understand or to solve more complicated problems.

Focusing on dialogue systems, while most current dialogue systems can treat only one domain (a small world for a single service), some research(Goddeau et al., 94; Namba et al., 94) which aims at increasing the domains, what is called a transportable system(Grosz, 83; Paris, 89) are now on-going. In such systems, information retrieval across multiple domains is realized using the relational databases. However in our system, it is difficult to retrieve information across multiple domains, because the information is retrieved from CD-ROMs in which a large amount of texts are contained, by using full-text retrieval techniques.

And while there are robust and useful strategies in certain goals, there isn't an all-powerful sterategy which covers all goals. If a robust strategy in a certain goal is introduced into the system, the user misunderstands that the system has an all-powerful strategy. Thus, in our system the user sometimes gets into trouble as follows:

- the user misunderstands that the information contained across several data sources can be obtained at once.
- the user is confused between a certain retrieval strategy which is robust in a certain goal and another simple but rather redundant strategy.

Furthermore, it is difficult to manage a discourse involving multiple goals in current dialogue systems. This is because most current systems aren't robust enough for anaphora and they are able to manage only a single and simple context. This sometimes causes the following problem:

- the user has to manage the multiple contexts involving multiple goals, because the system only manages a single context. And this makes it hard for the user to use the system.

As the result, the user also gets lost in the system.

In this paper, we focus on how to make the user aware of what the system can or cannot do. Thus, we propose a new dialogue system with multiple agents, in which we introduce the concept of multi-agent system into our dialogue system. In our system, three types of dialogue agents are realized: 1) for each domain, 2) for each strategy and 3) for each context. These agents take turns and play their roles according to the discourse situations. With these agents, our system will have the following characteristics:

- the domain agents make the user aware of the boundary between unintegrated domains.
- the strategy agents make the user aware of the difference between the domain oriented strategies.

- the context agents make it easy for the user to deal with the complicated discourse involving multiple goals.

In this paper, we first explain our baseline spoken dialogue system TARSAN which deal with multiple domains. Secondly, we describe the problems which arise when we extend the system into multiple domains. After that, we propose a new dialogue system with multiple dialogue agents. We also describe the results of the examinations on the proposed system. Finally, we conclude the paper.

2 The baseline system: TARSAN

We have been constructing a spoken dialogue system which retrieves information from a large amount of texts contained in CD-ROMs, named TARSAN(Sakai et al., 94; Sakai et al., 95). Figure 1 shows the configuration of the baseline system TARSAN for multiple domains.

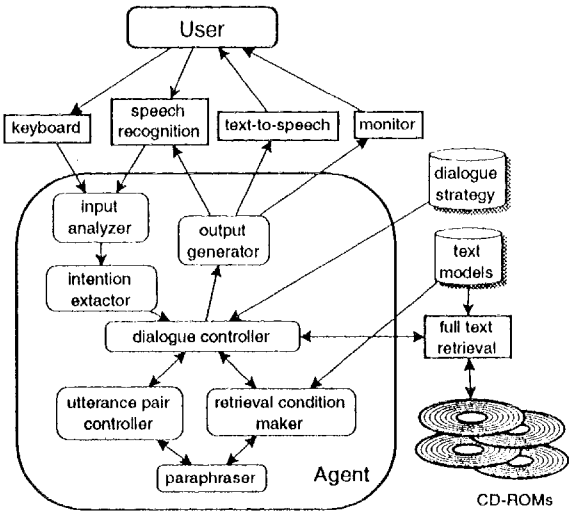

Figure 1: The configuration of TARSAN for multiple domains

TARSAN retrieves the information using the following processes:

1. The input analyzer analyzes the result of the speech recognition or the sentence received from keyboard.
2. The intention extractor extracts the user's intention (i.e. question, answer, condition change, and so on) based on the analysis of the modality.
3. The utterance pair controller deals with not only a simple pair of QA but also deals with follow-up questions based on utterance pair controlling.
4. The retrieval condition maker makes retrieval conditions which is sent to the full text retrieval process by the dialogue controller described below. The retrieval conditions are created by referring the 'text-models', which define the relation between the input words and the retrieval conditions.
5. The paraphraser translates various expressions of the inputs into a single domain oriented concept.
6. The dialogue controller determines the system's behavior (to retrieve and to answer the result, or to request more retrieval conditions to the user) by referring the retrieval conditions and the dialogue strategy.
7. The output generator generates the output sentence to be announced by the text-to-speech process and the information to be displayed on the monitor.

Our current system TARSAN is able to access the following four CD-ROMs:

CD-ROM1: sight-seeing information in Japan (i.e. name, location, explanation, and so on of temples, hot springs, golf courses, and so on)(Kosaido, 90).

CD-ROM2: hotel information in Japan (i.e. name, telephone number, room charges, equipment, and so on)(JTB, 92).

CD-ROM3: Japanese and foreign cinema information(i.e. title, cast, director, story, and so on)(PIA, 90).

CD-ROM4: Japanese professional baseball player information(i.e. name, belonging team, records, and so on)(Nichigai, 90).

TARSAN treats CD-ROM1 and 2 as a single travel domain, CD-ROM3 as a cinema domain, and CD-ROM4 as a baseball domain.

3 Problems

As we described in the introduction, we have addressed three main problems in our dialogue system. Two problems derive from the extension of the system to multiple domains. And the last one derives from the single path contextual management.

1. The first problem is that the user misunderstands that the information contained across several data sources can be obtained by a single input sentence. The following are examples of requests accross domains: The first example is contained in the cinema domain and in the travel domain, and the second example is contained in the baseball domain and in the cinema domain.

Example 1: "*Yamaguchi Momoe ga shuen sita eiga no butai ni natta onsen wo shiritai.*"
(I want to know the hot spring which is the scene of the cinema whose star is Yamaguchi Momoe.)

Example 2: "*Puro yakyuu senshu datta haiyuu ga shutsuen sita eiga wo oshiete.*"
(Tell me the cinema where an actor who was a professional baseball player performs.)

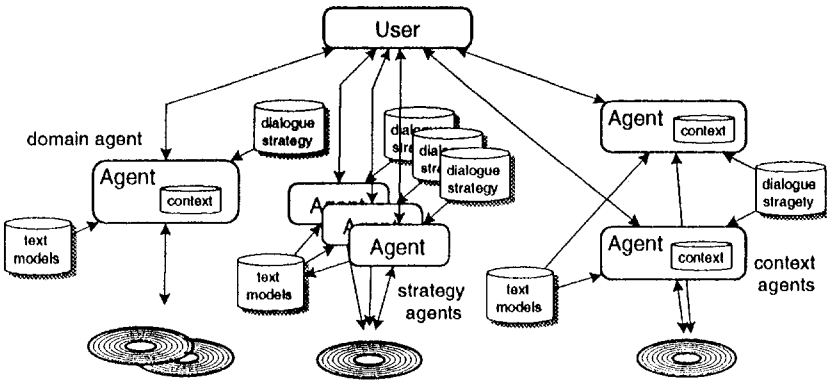

Figure 2: Three types of agents

2. The second problem is that the user misunderstands that the system has an all-powerful strategy, if it has a robust strategy for a certain purpose. Suppose that several discourse strategies exist in a single dialogue agent: one is a very sophisticated but very goal specific strategy which allows the user to reach the goal immediately, and another is a very simple but redundant strategy which has the ability to achieve any kind of goal. In this case, the user may confuse the potential of these strategies and feel uncomfortable about the gap.

3. The last problem is that the user has to manage multiple contexts concerning to multiple goals, because the system isn't enough robust for anaphora and only manages a single context. And this makes it hard for the user to use the system. Table 1 is an example that the user compares the information between Hakone and Nikko[1]. The example shows that the user has managed the context himself, which seems very complicated.

We have also assumed that these three problems arise because the system only has a single dialogue agent. A single dialogue agent usually deals with everything and this makes the user invisible what the system can or cannot do. Thus, we propose a new dialogue system with multiple agents which make the system's ability more visible to the user.

4 Dialogue system with multiple dialogue agents

In this section, we introduce a new dialogue system with multiple dialogue agents. The purpose is to make the user aware of what the system can or cannot do. In our system, three types of dialogue agents are realized: 1) for each domain, 2) for each strategy and 3) for the each context. Here, we call these agents as 1) domain agents, 2) strategy agents, 3) context agents, respectively. Figure 2 shows a brief sketch of these three types of agents. These agents take turns and play their roles according to the discourse situations. The details of these agents are as follows.

[1]They are famous sight-seeing places in Japan.

Table 1: An example that the user manage the multiple-goals by oneself

usr1:	*Hakone ni aru onsen wo oshiete.* (Tell me the hot springs in Hakone Town.)
sys1:	*16 ken arimasu.* (There are 16 hot springs.)
usr2:	*Nikko deha.* (How about in Nikko?)
sys2:	*Chuuzenji onsen, Nikko yumoto onsen ga arimasu.* (There are Chuuzenji onsen and Nikko yumoto onsen.)
usr3:	*Hakone niha jiin ga arimasuka.* (Are there any temples in Hakone?)
sys3:	*Amida dera, Kuduryu Myojin, Saunji nado 7 ken arimasu.* (There are 7 temples; Amida dera, Kuduryu Myojin, Saunji, and so on.)
usr4:	*Nikko niha.* (How about in Nikko?)
sys4:	*Nikko Toshoguu ga arimasu.* (There is Nikko Toshoguu.)
usr5:	*Sono setsumei wo kikitai.* (Please explain it.)
sys5:	*Tokugawa Ieyasu no rei wo matsuru. ...* (The soul of Tokugawa Ieyasu is worshipped. ...)

4.1 The domain agents

To solve the first problem, we realized domain agents which perform information retrieval in each different domain. Figure 3 shows a brief sketch of the domain agents. The domain agents perform the basic interaction between the user and the system to retrieve the information in the basic manner specific to each domain. In every domain agent, indispensable and basic conditions for information retrieval are defined. Using these conditions, the domain agent communicates with the user and performs the information retrieval. And when the user's input moves from one domain to another domain, the domain agent will also change. Thus with the domain agents, the user is made aware of the boundary between the domains. We expect this mechanism to prevent the user from asking the question across uninte-

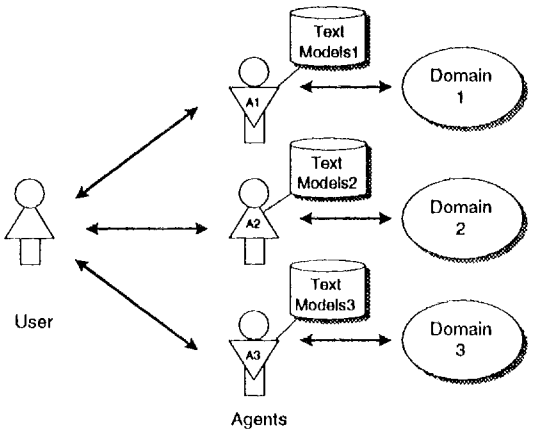

Figure 3: The domain agents

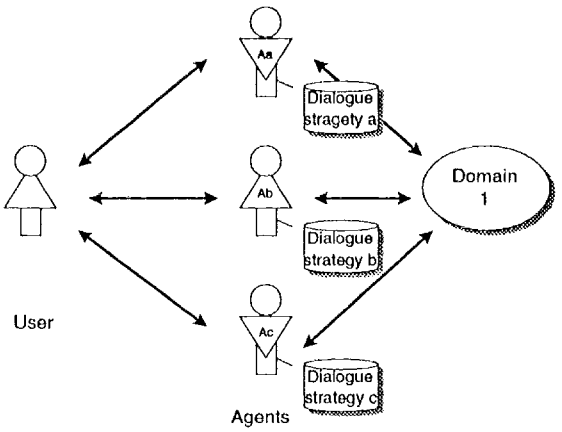

Figure 4: The strategy agents

Table 2: An example of two agents try to make an action

usr:	"*Yamaguchi Momoe ga shuen sita eiga no butai ni natta onsen wo shiritai.*" (I want to know the hot spring which is the scene of the cinema whose main cast is Yamaguchi Momoe.)
C.agt:	"*Izu no odoriko, Shunkinsho, nado 13 ken arimasu.*" (There are 13 cinemas: Izu no Odoriko, Shunkinsho, and so on)
T.agt:	"*Jouken ni gaitousuru onsen ha arimasen.*" (There is no hot spring satisfying the condition.)

grated multiple domains. For example, in the case of the example 1 in section 3, two agents dealing with the cinema domain and the travel domain try to make each action as Table 2 shows[2]. Thus the user will be aware of the boundary between the two domains.

4.2 The strategy agents

To solve the second problem, we realized the strategy agents which performs information retrieval according to each specific strategy for the information retrieval. Figure 4 shows a brief sketch of the strategy agents. The strategy agents handle the interaction between the user and the system to retrieve the information in the manner specific to each task. In every strategy agent, task specific conditions for the information retrieval are defined. Using the task specific conditions, the strategy agent is able to use the default condition specific to the task and is able to give advice or to give choices to the user. Thus with the strategy agents, the user is made aware of the strategy which is specific to the task and this mechanism prevents the user using the task specific strategy for other tasks.

In the current system, there are two strategy agents for the travel domain:

business trip strategy agent: indispensable condition for the input is the destination, and the optional conditions are the room charge and the circumstances. When the optional conditions are not defined by the user, the strategy agent will recommend some choices to the user. The default responses are the name of the hotel and its telephone number in this task.

recreation strategy agent: indispensable condition for the input is the recreation equipment and the number of participants and the other conditions are optional. When the optional conditions are not defined by the user, the strategy agent will recommend some choices to the user. The default responses are also the name of the hotel and its telephone number in this task.

These strategy agents not only allow the user to use the system easily but also help the user to be aware of the characteristics of the dialogue strategy specific to the task.

Table 3 compares the difference between using the domain agent for travel and the business trip strategy agent. As you can see from the table, more friendly discourse is achieved when using the strategy agent.

4.3 The context agents

To solve the last problem, we realized the context agents which perform the information retrieval dependent in different contexts. The context agent is defined when the user moves from one context to another. Figure 5 shows a brief sketch of the context agents. Using the context agents the user can easily compare the results relating to multiple goals.

Table 4 is the dialogue which aims at the same goal as table 1 by using multiple context agents. As you can see from the table 1 and 4, more simplified discourse is achieved when using context agents.

[2]Travel agent is able to retrive and find "the hot spring which is the scene of Izu no odoriko".

Table 3: The comparison between the discourse using the domain agent and the strategy agent

domain agent for travel	
agt0:	*Hai, Ryokou eejent desu.* (Hello, I'm travel agent.)
usr1:	*Tottori shi no hoteru wo shiritai.* (I want to know the hotels in Tottori city.)
agt1:	*34 ken arimasu.* (34 rows exist.)
usr2:	*Ekimae ni aru hoteru.* (Land circumstance is near the station.)
agt2:	*8 ken arimasu.* (8 rows exist.)
usr3:	*shukuhaku ryou ga 8000yen ika deha.* (And room charge is under 8000yen.)
agt3:	*Hoka no jouken "Ricchi jouken ga ekimae" wo nokoshi masuka.* (Do you keep another condtion, "Land circumstance is near the station"?)
usr4:	*Hai.* (Yes.)
agt4:	*AAA hoteru, hoteru BBB desu.* (AAA hotel and hotel BBB exist.)
usr5:	*Sono denwabangou wo oshiete.* (Tell me the telephone numbers.)
agt5:	*AAA hoteru no denwabangou ha XXX-XXXX, hoteru BBB no denwabanou ha YYY-YYYY desu.* (The telephone number of AAA hotel is XXX-XXXX, and the one of hotel BBB is YYY-YYYY.)

business trip strategy agent	
agt0:	*Hai, Shucchou eejento desu. Dochira ni shucchou nasai masu ka.* (Hello, I'm business trip agent. Where are you goingo to go?)
usr1:	*Tottori shi desu.* (Tottori city.)
agt1:	*Ricchi jouken ha ekimae, shukuhaku ryou ha 8000yen ika de yoroshii desuka.* (Circumstance is near the station, and room charge is under 8000yen. Are these OK?)
usr2:	*Hai.* (Yes.)
agt2:	*2ken arimasu. AAA hoteru no denwabangou ha XXX-XXXX, hoteru BBB no denwabanou ha YYY-YYYY desu.* (2 hotels exist. The telephone number of AAA hotel is XXX-XXXX, and the one of hotel BBB is YYY-YYYY.)

5 Examinations

In this section, we described the examinations of the proposed system. In order to examine the effectiveness of the multiple dialogue agent system (new system), we compare it with the single dialogue agent system (old system). Here the single dialogue agent is the domain agent for the travel domain.

We evaluated the system by counting the number of the interactions between the user and the system (Turns), the number of input characters of the users (Characters), and session time (Seconds) that subjects took to reach the same goal with new system and the old one. Eight subjects examined these systems. They are all typists, but novices with dialogue systems. They were given a brief explanation of both systems and practiced on them for about quarter an hour each. We divide

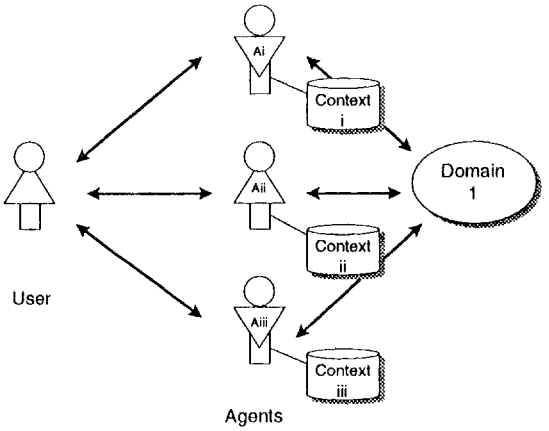

Figure 5: The context agents

Table 4: The dialogue using two context agents

(to Hakone agent and Nikko agent)	
usr1:	*Onsen wo shiritai.* (I want to know the hot springs.)
H.agt1:	*16 ken arimasu.* (There are 16 hot springs.)
N.agt1:	*Chuuzenji onsen, Nikko yumoto onsen ga arimasu.* (There are Chuuzenji onsen and Nikko yumoto onsen.)
(to both agents)	
usr2:	*Jiin ha arimasuka* (Are there any temples?)
H.agt2:	*Amida dera, Kuduryu Myojin, Saunji nado 7 ken arimasu.* (There are 7 temples; Amida dera, Kuduryu Myojin, Saunji, and so on.)
N.agt2:	*Nikko Toshoguu ga arimasu.* (There is Nikko Toshoguu.)
(to Nikko agent)	
usr3:	*Sono setsumei wo kikitai.* (Please explain it.)
N.agt3:	*Tokugawa Ieyasu no rei wo matsuru. ...* (The soul of Tokugawa Ieyasu is worshipped. ...)

the subjects into two groups. Group 1 examined new system first and old one next, and group 2 did old system first and new one next.

5.1 Examination 1

The following goal is given to every subject:

Goal 1: You will go to Kurashiki City on business. Find a suitable hotel[3]. (You may select different hotels with each system.)

The relevant agent in the new system is the business trip agent. Table 5 shows the results (averages of Turns, Characters, and Seconds) of examination 1. These results show not only that both groups needed less dialogue using new system than using old system, but also that group 1 needed less dialogue, especially less session time (360:640), when they used old system than group 2. This

[3]There are 41 hotels in Kurashiki City.

means that the user is able to learn how to use the old (strategy-less) system by using new system with a typical strategy. We also mention that all six subjects who selected different hotels were happy about the hotel using the new system.

Table 5: The results of examination 1

	old → new		new → old	
Turns	7.3	3.0	3.5	5.5
Characters	75	18	25	56
Seconds	640	175	190	360

5.2 Examination 2

The following goal is given to every subject:

Goal 2: You have to select Kanazawa or Sendai for sight-seeing. Compare them using some retrieved information, and select one.

The relevant agents in the new system are Kanazawa agent and Sendai agent. Table 6 is the results of examination 2. These results show an interesting phenomenon that in the case of the dialogue comparing multiple goals with these complicated processes, the user tends to stop comparing by session time (from five minutes to ten minites) in favour of the obtained retrieval results. And the new system is able to obtain more retrieval results than the old system. Thus the new system is better than the old system in the case of dealing with multiple goals.

Table 6: The results of examination 2

	old → new		new → old	
Turns	7.0	10.3	9.3	8.5
Characters	79	54	51	96
Seconds	442	420	458	526

6 Conclusion

In this paper, we proposed a new dialogue system with multiple dialogue agents. In our new system, three types of agents were realized. They were a) domain agents, b) strategy agents, and c) context agents. These agents give the following advantages to the user:

- the domain agents prevent the user from asking the questions across unintegrated domains.
- the strategy agents make the user aware of the difference between the domain oriented strategies.
- the context agents make it easy for the user to deal with the complicated discourse involving multiple goals.

Using these agents, we expect the user to understand what the system can or cannot do. The experimental results show that the user can retrieve effectively and obtain the expected goals easily by using these multiple agents.

Acknowledgement

The authors wish to thank Dr. Hideyuki Tamura, head of the Media Technology Lab., for giving the opportunity of this study, Dr. Yasuhiro Komori and Tom Wachtel for suitable advice in translating this paper into English, and members of the Intelligent Media Div. for useful discussions.

References

Cécile L.Paris. 1989. The Use of Explicit User Models in a Generation System for Tailoring Answers to the User's Level of Expertise. User Models in Dialog Systems, pp.200–pp.232. Springer-Verlag, A.Kobsa, W.Wahler (Eds.).

Goddeau D., Brill E., Glass J., Pao C., Phillips M, Polifroni J., Seneff S., and Zue V. 1994. GALAXY: A Human-language Interface to Online Travel Information. In *Proceedings of ICSLP-94*, S13-11.

Grosz B.J. 1983. A Transportable Natural-Language Interface System. In *Proceedings of Conference on Applied Natural Language Processing*, pp.25–pp.30.

JTB. 1992. *JTB's Accommodation Information* (Electronic Book YRRS-094). ed. by Nakajime R. JTB. Tokyo.

Kosaido. 1990. *TABIGURA* (Electronic Book YRRS-028). ed. by Nihon-Kanko-Kyokai Corporation. Tokyo.

Maes P. 1991. The Agent Network Architecture (ANA). SIGART bulletin (2), 4.

Nagao K. and Takeuchi A. 1994. Social Interaction: Multimodal Conversation with Social Agents. In *Proceedings of AAAI-94*, pp. 22–28.

Namba Y., Tsuji H., and Kinugawa H. 1994. Natural Language Interface for Multiple Systems Sequential Control (In Japanese). IPSJ Vol.35 No.1, pp.20–34.

Nichigai. 1990. *The professional baseball* (Electronic Book YRRS-023). Morioka H. and Nichigai Asscociates Co. Tokyo.

Nishida T. and Takeda H. 1993. Towards the Knowledgeable Community. In *Proceedings of International Conference on Building and Sharing of Very-Large Scale Knowledge Bases '93 (KB & KS '93)*, pp. 157–166. Ohmsha.

PIA. 1990. *PIA cinema club* (Electronic Book YRRS-016). ed. by Watanabe K. PIA Co.

Sakai K., Ikeda Y., and Fujita M. 1994. Robust Discourse Processing Considering Misrecognition in Spoken Dialogue System. In *Proceedings of ICSLP-94*, S17-7.

Sakai K., Ikeda Y., and Fujita M. 1995. *Gaidansu Sisutemu no Maruchi Domein-ka no Kokoromi* (In Japanese), (A trial to Increase Domains in Guidance System). In *Proceedings of The First Annual Meeting of The Association for Natural Language Processing*, pp. 301–304.