A Corpus-based Approach for Cooperative Response Generation in a Dialog System

Zhiyong Wu, Helen Meng, Hui Ning, and Sam C. Tse

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Shatin
john.zy.wu@gmail.com, {hmmeng, hning, cftse}@se.cuhk.edu.hk

Abstract. This paper presents a corpus-based approach for cooperative response generation in a spoken dialog system for the Hong Kong tourism domain. A corpus with 3874 requests and responses is collected using Wizard-of-Oz framework. The corpus then undergoes a regularization process that simplifies the interactions to ease subsequent modeling. A semi-automatic process is developed to annotate each utterance in the dialog turns in terms of their key concepts (KC), task goal (TG) and dialog acts (DA). TG and DA characterize the informational goal and communicative goal of the utterance respectively. The annotation procedure is integrated with a dialog modeling heuristic and a discourse inheritance strategy to generate a semantic abstraction (SA), in the form of $\{TG, DA, KC\}$, for each user request and system response in the dialog. Semantic transitions, i.e. $\{TG, DA, KC\}_{\text{user}} \rightarrow \{TG, DA, KC\}_{\text{system}}$, may hence be directly derived from the corpus as rules for *response message planning*. Related verbalization methods may also be derived from the corpus and used as templates for *response message realization*. All the rules and templates are stored externally in a human-readable text file which brings the advantage of easy extensibility of the system. Evaluation of this corpus based approach shows that 83% of the generated responses are coherent with the user's request and qualitative rating achieves a score of 4.0 on a five-point Likert scale.

Keywords: Natural language generation (NLG), Response generation, Corpus-based approach

1 Introduction

Continual advancements in speech and language technologies have brought usable spoken dialog systems (SDS) within reach. SDS typically supports goal-oriented human-computer conversations regarding restricted application domains, e.g. asking for a restaurant recommendation, planning a trip, etc. SDS integrates technologies including speech recognition (SR), natural language understanding (NLU), dialog modeling, information/database access and text-to-speech synthesis. An indispensable component that facilitates effective two-way, human-computer interaction is natural language generation (NLG) of *cooperative system responses* that tailor to the user's information needs and linguistic preferences. NLG is defined as the process of transforming a semantic specification from the dialog model (DM) into a *semantically* *well-posed* and *syntactically well-formed* message. The message can be presented to the user as on-screen text and/or synthesized speech. The demarcation between the DM and NLG may vary from one system to another. Some (earlier) systems do not distinguish between the two processes. In this work, the demarcation is drawn whereby the DM provides discourse-inherited semantics for the NLG. The NLG aims to compose a well-posed and well-formed message that can serve as a cooperative system response. To compose a well-posed message, the NLG needs to select content pertinent to the current dialog turn and cast the content in a message plan that is coherent and succinct. To compose a well-formed message, the NLG needs to select syntactic and elements for textual/audio realization of the response message. We divide NLG problem into two sub-problems – (i) *message planning* formulates a well-posed message plan (MP) based on relevant semantics; and (ii) *message realization* generates a well-formed linguistic realization from the MP.

Previous approaches in NLG generally fall within a continuum between the non-linguistic template-based approach and the fully linguistic approach [1,2]. The template-based approach has been widely adopted due to ease of development, maintenance and predictability [3]. However, handcrafting templates for every application domain is a tedious process with low portability. It is often impossible to handcraft templates that fully cover the combinatoric space of communicative goals and discourse contexts. Hence templates offer limited variety and the approach becomes untenable as the application domain grows. Fully linguistic approaches mostly originate from research in NLG of monologs (e.g. reports, summaries, etc.) and incorporate a huge amount of linguistic knowledge [4,5]. Adapting these approaches for dialogs in restricted domains and achieving real-time performance may be difficult [6]. Recent efforts in NLG research strive to strike a balance between the non-linguistic and fully linguistic ends of the spectrum, by using simple rules/grammars augmented with corpus-based statistics. This can reduce the need for a full linguistic characterization and can also introduce variety into the NLG output [7]. A representative example is the use of stochastically combined dialog acts to form surface realizations and these are then selected by a filter trained on a human-graded corpus [8].

In this paper, we present an approach where message planning strategies and message realization templates are derived from a dialog corpus. This approach can vastly reduce the human effort that needs to be devoted to authoring rules and templates. We collected a dialog corpus by means of a Wizard-of-Oz setup, where the "wizard" attempts with best effort to answer to the user's inquiries in a systematic and succinct way. The collected data then undergoes a manual "regularization" process for simplification in order to ease subsequent modeling. We also designed a semantic abstraction of each user's request and system's response, in terms of key concepts, tasks goals (i.e. the informational goals) and dialog acts (i.e. communicative goals). Hence we may capture the message planning strategies found in the corpus through semantic transitions of a pair of request/response turns. For a given message plan, we may also refer to the corpus to derive message realization templates. This corpus-based approach eases development of the NLG component and may enhance portability across languages and applications. In the following, we present the details of corpus development, semantic abstraction and annotation, message planning, message realization as well as evaluation results.

2 Corpus Development

2.1 Information Domain

The information domain is specific to Hong Kong tourism, as defined by the Tourism Board's website – Discover Hong Kong.[1] The domain covers information ranging from scenic attractions, shopping attractions, transportation, fare prices, events, tours, etc. This diversity is useful for our current research in natural language generation.

Based on the website, we also developed a database covering 349 attractions. Related information constituents that are tagged with XML (eXtensible Markup Language) include name, type, description, routing, time, url, etc. An example is shown in Table 1 for illustration.

Table 1. An example of XML-tagged data entry in the Hong Kong tourism domain.

```
<ATTRACTION>
   <NAME>迪士尼樂園</NAME>   (translation: Disneyland)
   <TYPE>主題公園</TYPE>   (theme park)
   <DESCRIPTION>
   從踏進香港迪士尼樂園那一刻開始，令人興奮着迷的奇妙之旅便已展開！……
   </DESCRIPTION>   (the Hong Kong Disneyland is an exciting place...)
   <ROUTE>在地鐵欣澳站轉乘迪士尼綫列車</ROUTE>   (take the mass transit railway
   (MTR) to Sunny Bay station and transit to the Disney line)
   <TIME>開放時間為上午10時至晚上8時</TIME>   (opening hours...)
   <PRICE>成人295，小童210，長者170</PRICE>   (fares for adults, children and seniors)
   <URL>http://hongkongdisneyland.com</URL>
</ATTRACTION>
```

2.2 Eliciting Interactions using a Wizard-of-Oz Data Collection

In order to elicit interactions in the selected domain, we use a Wizard-of-Oz (WoZ) data collection setup to elicit interactions from a group of thirty invited subjects. Each subject and the wizard sat in different rooms, and interacted through a multimodal and multimedia interface through networked computers. The subjects can issue inquiries using speech, typed text and/or pen gestures. The wizard can refer to the Discover Hong Kong website during the entire data collection process and always tries to respond to the user's inquiries with best effort. All interactions were logged by the system. As a result of this data collection process, we have a series of dialogs that contain rather free-form wizard-generated responses for the subjects' inquiries. These enable us to define the major informational goals (or task goals, TG) of the subjects, dialog acts (DA) that characterize the course of dialog interactions and related key concepts (KC) that may come from the current interaction (dialog turn) or inherited from previous interactions.

2.3 Data Regularization Process

[1] http://www.discoverhongkong.com

As mentioned earlier, the wizard's responses as logged from the WoZ data collection procedure is relatively free form. It contains many disfluencies such as filled pause, word order reversal due to spontaneity in interaction and tagged information indicating responses in alternative modalities, e.g. highlighted points on a map, urls, etc. In order to ease the subsequent process of modeling the dialog responses, we devised a manual procedure of data regularization where the collected data are simplified into short sentences/utterances with straightforward structures. This paves the way of easing the development of message templates for verbalization of relevant information content. In total, we have regularized the entire dialog corpus, which consists of 1500 dialog turns, each with two to five utterances. Overall, there are 3874 request and response utterances. Table 2 shows a simple dialog interaction before and after the data regularization process.

Table 2. An example dialog between the Wizard (W) and the User (U) before (Original data) and after (Refined data) the data regularization process.

	Original data	Refined data
W0	請問，你第二天想去哪裏？	請問，你第二天想去哪裏？ *(Where would you like to go on the second day?)*
U1	Er，主題公園…我想去看看。	我想去主題公園看看。*(I would like to visit theme park)*
W1	請問你想去海洋公園，還是迪士尼樂園？	請問你想去海洋公園，還是迪士尼樂園？ *(Would you like to visit Ocean Park or Disneyland?)*
U2	Um，讓我想想。去迪士尼樂園好了。	讓我想想。去迪士尼樂園好了。 *(Let me think. I prefer Disneyland.)*
W2	這裏是迪士尼樂園的資料<url>，請看。	這裏是迪士尼樂園的資料，請看。*(Here is the information about Disneyland, please have a look.)*
U3	從中環到這裏<point: pictures>怎麼走呢？	從中環到這裏怎麼走呢？ *(How could I get here from Central?)*
W3	Er，從中環到迪士尼樂園的話，你可以在欣澳站轉乘地鐵迪士尼綫列車就到了。	從中環到迪士尼樂園，你可以在地鐵欣澳站轉乘迪士尼綫列車。*(From Central to Disneyland, you can take the MTR to Sunny Bay station and transit to the Disney line.)*
U4	那麼，有沒有那個Er海洋公園的資料？	有沒有海洋公園的資料？ *(Is there any introduction about Ocean Park?)*
W4	這個就是海洋公園的資料<url>，請看。	這個就是海洋公園的資料，請看。*(This is the information about Ocean Park, please have a look.)*
U5	再見。	再見。*(Bye-bye.)*
W5	祝你旅途愉快！	祝你旅途愉快！ *(Have a good trip!)*

3 Semi-Automatic Corpus Annotation of Semantic Constituents

A critical stage in corpus development is the annotation of major semantic constituents in the collected data. These semantic constituents must characterize: (i) what are the types of questions asked; (ii) what kinds of content are necessary for answering these questions (i.e. *response message planning*); and (iii) how such content should be expressed (i.e. *response message realization*). As mentioned above, we believe that the major semantic constituents needed include the key concepts (KC) in a verbal message; the domain-specific task goal (TG) underlying the message; as well as the communicative role of the message in the course of the dialog, as symbolized by the dialog act (DA) [9,10]. We have devised a semi-automatic method of annotating such semantic constituents. The objective is to reduce the manual effort needed, speed up the annotation process, as well as to enhance consistency in the annotations.

3.1 Tagging Key Concepts (KC)

We defined approximate 800 grammar rules (in the form of regular expressions) from analyzing the collected data for tagging concepts. Examples are shown in Table 3.

Table 3. Example of grammar rules for tagging key concepts (KC).

attraction → 迪士尼樂園 | 海洋公園 | 主題公園 | …	*(Disneyland | Ocean park | Theme park |..)*
how → 怎麼 | 如何 | …	*(These Chinese tokens mean "how to")*
go → 走 | 行 …	*(These Chinese tokens mean "go or walk")*
origin → 從 [attraction]	*(from [attraction])*
destination → 到 [attraction]	*(to [attraction])*
directions → [how] [go]	

3.2 Task Goals and Dialog Acts

The task goal (TG) symbolizes the information goal of the user's request and is domain-specific. The dialog act (DA) expresses the communicative goal of an expression in the course of a dialog and bears relationships with the neighboring dialog turns. The DA is largely domain-independent. We defined 12 TGs based on the collected corpus, as shown in Table 4. We also included 17 DAs, adapted from VERBMOBIL-2 [9], as shown in Table 5.

Table 4. 12 Hong Kong tourism domain specific task goals (TGs).

ATTRACTION, DURATION, FEE, LOCATION, PHONE, ROUTE, SHOPING, HOURS, TOURING, FOOD, HOTEL, RESERVATIONS

Table 5. 17 domain independent dialog acts (DAs).

APOLOGY, BYE, BACKCHANNEL, CLOSE, CONFIRM, DEFER, GREET, SUGGEST, THANK, FEEDBACK_NEGATIVE, FEEDBACK_POSITIVE, REQUEST_SUGGEST, REQUEST_COMMENT, REQUEST_DETAILS, REQUEST_PREFERENCE, INFORM_GENERAL, INFORM_DETAILS

3.3 Semi-Automatic Annotation Process

Each dialog turn in the regularized corpus is segmented into individual utterances such that each utterance corresponds to only one TG and one DA. For example, the second user's request (U2) "讓我想想。去迪士尼樂園好了。*(Let me think. I prefer Disneyland.)*" (see Table 2) is segmented into two utterances as shown in Table 7 – "讓我想想 *(Let me think.)*" followed by "去迪士尼樂園好了 *(I prefer Disneyland.)*".

We divided the corpus into four subsets, as shown in Table 6. The annotation procedure is incremental. We first hand-annotate data subset #1 in terms of TG and DA. KCs are tagged by the regular expressions mentioned above. All annotations (KC, TG and DA) are checked by hand and are used to train a suite of Belief Networks (BNs) [11] that can accept a series of input KCs from an utterance and output the TG and DA labels for the utterance. These BNs are used to label data subset #2 which then undergoes a pass of manual checking. Thereafter, both data subsets #1 and #2 are used to retrain the BNs and these are subsequently used to label data subset #3, which again undergoes a pass of manual checking. Thereafter, all three data subsets are used to retrain the BNs and these are evaluated based on data subset #4. The BNs achieve an accuracy of 79% for TG and 77% for DA labeling in data subset #4. Table 7 illustrates an example of the end result of this annotation process. Every utterance in a dialog turn may thus be annotated with KCs, TG and DA.

Table 6. Division of the Corpus into four data subsets.

Data subset	#1	#2	#3	#4
Number of utterances	948	939	986	1001

Table 7. Results of annotation, based on the example presented earlier in Table 2.

W0	請問，你第二天想去哪裏？ KC: {ask_where=去哪裏} TG: ATTRACTION	 DA: REQUEST_PREFERENCE
U1	我想去主題公園看看。 KC: {attraction=主題公園} TG: ATTRACTION	 DA: INFORM_DETAILS
W1	請問你想去海洋公園，還是迪士尼樂園？ KC: {attraction=海洋公園, attraction=迪士尼樂園} TG: ATTRACTION	 DA: REQUEST_COMMENT
U2	讓我想想。 KC: {think=想想} TG: ATTRACTION	 DA: DEFER
U2	去迪士尼樂園好了。 KC: {attraction=迪士尼樂園} TG: ATTRACTION	 DA: INFORM_DETAILS
W2	這裏是迪士尼樂園的資料，請看。 KC: {attraction=迪士尼樂園} TG: ATTRACTION	 DA: INFORM_GENERAL
U3	從中環到這裏怎麼走呢？ KC: {origin=中環, destination=這裏, directions=怎麼走} TG: ROUTE	 DA: REQUEST_DETAILS
W3	從中環到迪士尼樂園，你可以在地鐵欣澳站轉乘迪士尼綫列車。 KC: {origin=中環, destination=迪士尼樂園, route=在地鐵..} TG: ROUTE	 DA: INFORM_DETAILS
U4	有沒有海洋公園的資料？ KC: {attraction=海洋公園} TG: ATTRACTION	 DA: REQUEST_DETAILS
W4	這裏就是海洋公園的資料，請看。 KC: {attraction=海洋公園} TG: ATTRACTION	 DA: INFORM_GENERAL

U5	再見。 KC: {bye=再見} TG: ATTRACTION DA: BYE
W5	祝你旅途愉快！ KC: {good_trip=旅途愉快} TG: ATTRACTION DA: CLOSE

4 Message Planning and Realization in Response Generation

The annotation procedure described in the previous section is applied to every user's request and system (wizard) response in the regularized corpus. In addition, our dialog model (DM) incorporates a heuristic that the TG of the ensuing response is assumed to be identical to the user's request since the system (wizard) is generating cooperative responses. The DM also incorporates a *selective discourse inheritance strategy* [12] to enhance the completeness of the semantic representation of an utterance. For example, if the user first asks "海洋公園怎麼去？" *(How can I get to Ocean Park?)*, followed by "迪士尼樂園呢？" *(How about Disneyland?)*, the second question must inherit appropriate concepts from the previous question in order to have a self-complete meaning. We have developed a set of context-dependent inheritance rules [12] that govern the inheritance of TG or KC from previous dialog turns. The extensions of the heuristic and discourse inheritance raised the TG and DA label accuracies to over 90% in data subset #4. Sequential processing by the semi-automatic annotation process and the DM transforms every user request and system (wizard) response in the collected corpus into a *succinct semantic representation*, in terms of $\{TG, DA, KC\}$. Such semantic abstraction (SA) of user's request and system's responses are useful for deriving strategies of response message planning as well as methods of response message realization. We will describe the two procedures in the following.

4.1 Strategies for Response Message Planning

Parsing for the TG, DA and KC in a regularized user request or system (wizard) response message automatically generates a semantic abstraction (SA) representation $\{TG, DA, KC\}$. Pairing up the SAs of a user's request with its system response in the subsequent dialog turn automatically derives message planning strategies in the form of semantic transitions, i.e.:

$$\{TG, DA, KC\}_{\text{user}} \rightarrow \{TG, DA, KC\}_{\text{system}}.$$

In other words, each pair of user-system interactions in the 3,874 utterances in our corpus offer one instance of message planning by the wizard in the context of the dialog system. Hence, our strategies for message planning are automatically derived in a data-driven manner.

We analyzed these instances and noted several features:

(i) A user's dialog turn may contain multiple utterances and each has its own SA representation. In such situations, the semantic transition rule is based only on the last utterance and its SA representation. This is because our corpus suggests that the last utterance can fully characterize the user's dialog turn. For example, the second user turn in Table 7 will only derive the SA representation of {*ATTRACTION*, *INFORM_DETAILS*, *attraction*}$_{\text{user}}$.

(ii) It is possible for different $\{TG, DA, KC\}_{\text{user}}$ to transit to the same $\{TG, DA, KC\}_{\text{system}}$. For example, in Table 7 the pair of dialog turns (U2, W2) produces {*ATTRACTION*, *INFORM_DETAILS*, *attraction*}$_{\text{user}}$→ {*ATTRACTION*, *INFORM_GENERAL*, *attraction*}$_{\text{system}}$; while the pair of dialog turns (U4, W4) produces {*ATTRACTION*, *REQUEST_DETAILS*, *attraction*}$_{\text{user}}$ → {*ATTRACTION*, *INFORM_GENERAL*, *attraction*}$_{\text{system}}$.

(iii) It is also possible for a given $\{TG, DA, KC\}_{\text{user}}$ to transit to several possible $\{TG, DA, KC\}_{\text{system}}$. For example, in Table 7, the pair of dialog turns (U1, W1) produces {*ATTRACTION*, *INFORM_DETAILS*, *attraction*}$_{\text{user}}$ → {*ATTRACTION*, *REQUEST_COMMENT*, *attraction*}$_{\text{system}}$. However, the pair of dialog turns (U2, W2) produces {*ATTRACTION*, *INFORM_DETAILS*, *attraction*}$_{\text{user}}$ → {*ATTRACTION*, *INFORM_GENERAL*, *attraction*}$_{\text{system}}$. This presents the need for devising a set of *rule selection conditions* in message planning. An illustration is presented in Table 8, where Rules 1 to 4 are all possible transitions originating from the same $\{TG, DA, KC\}_{\text{user}}$. It should be noted that these rule selection conditions are inserted manually upon analysis of the corpus. However, as illustrated in Table 8, these simple conditions should be generalizable to other information domains.

Table 8. An example of semantic transition rules which constitutes the message planning strategies for cooperative response generation. Rule selection conditions may be applied if there are multiple possible message plan options. These conditions may contain key concepts (denoted by '#') whose values are obtained either from database retrieval results (denoted by *database#concept*) or from the parsed user request (denoted by *request#concept*).

Semantic Transition Rule Format {TG, DA, KC}user→{TG, DA, KC}system
Rule 1
{ATTRACTION, INFORM_DETAILS, attraction}user→
{ATTRACTION, REQUEST_COMMENT, place}system
Rule 2
{ATTRACTION, INFORM_DETAILS, attraction}user→
{ATTRACTION, INFORM_GENERAL, attraction}system
Rule 3
{ATTRACTION, INFORM_DETAILS, attraction}user→
{ATTRACTION, INFORM_DETAILS, attraction}system
Rule 4
{ATTRACTION, INFORM_DETAILS, attraction}user→
{ATTRACTION, APOLOGY, sorry}system
Control Conditions for the above Rules Selection
IF ({database#result_number}>1) THEN select Rule 1
ELSEIF ({database#result_number}==0) THEN select Rule 4
ELSEIF ({request#detail}!=null) THEN select Rule 3
ELSEIF ({database#url}!=null) || ({database#picture}!=null) THEN select Rule 2

The introduction of *rule selection conditions* adds context-dependent variability in cooperative response generation. Referring to the dialog in Table 7 and the conditions in Table 8, the various conditions are:

- For the user request U1 in Table 7, the system finds several matching attractions related to the concept "attraction=主題公園 *(theme park)*" in the database. Hence the first rule selection condition {database#result_number}>1 in Table 8 is satisfied and Rule 1 is used as the message plan. Hence the system presents all matching options to the user in the generated response (W1) and seeks the user's input by the dialog act REQUEST_COMMENT.
- The user's feedback in U2 of Table 7 sets the concept value of "attraction=迪士尼樂園 *(Disneyland)*". The system can only find one matching entry in the database which comes with URL information. Hence the fourth rule selection condition in Table 8 {database#url}!=null is satisfied and Rule 2 is used as the message plan. This generates the response W2 under the dialog act of INFORM_GENERAL.
- If the user were to follow up with an utterance such as "給我介紹迪士尼樂園的詳細資料" *(Give me more details about Disneyland)*, which sets the concept value "detail=詳細 *(details)*", then the third rule condition in Table 8 is satisfied and Rule 3 will be used as the message plan.
- If the user requested an attraction which cannot be found in the database, then the second rule selection condition in Table 8 is satisfied and Rule 4 will be selected as the message plan. As a consequence, the system will apologize for not being able to offer relevant information.

4.2 Response Message Realization using Corpus-derived Templates

The semantic transitions above generates a message plan for generating the system response, in the form of semantic abstraction (SA) $\{TG, DA, KC\}_{\text{system}}$. Analysis of our regularized corpus also suggests that each of these SA may be verbalized in a variety of ways. These verbalization methods found in the corpus are encoded in a set of 89 message realization templates with labels, e.g. GENERAL_INFO, PICTURE_INFO, GOOD_TRIP, etc., as shown in Table 9.

Table 9. Examples of message realization templates derived from the regularized corpus.

Text Generation Templates:

Template Label: GENERAL_INFO
Contents: 這裏是{request#attraction}的資料，請看。
(translation: here is the information about {request#attraction}).

Template Label: PICTURE_INFO
Contents: 我想你可以看看{request#attraction}的圖片資料{database#picture}
(translation: you may refer to these pictures {database#picture} of {request#attraction}).

Template Label: GOOD_TRIP
Contents: 祝你旅途愉快！
(translation: have a good trip)

A given $\{TG, DA, KC\}_{\text{system}}$ may correspond to one or more message realization templates. In cases where there are multiple options, we devise a set of template selection rule based on the regularized corpus. This is illustrated in Table 10, where the system response with SA *{ATTRACTİON, İNFORM_GENERAL, attraction}*$_{system}$ may be verbalized by the templates GENERAL_İNFO or PİCTURE_İNFO, depending on whether database retrieval can provide a picture, i.e. {database#picture}!=null). All system reponses with the dialog act of BYE, regardless of the task goal, will be realized by the template GOOD_TRIP.

Table 10. Illustration of a template selection rule among possible message realization templates that correspond to a given $\{TG, DA, KC\}_{\text{system}}$. The asterisk (*) is a wildcard that matches all task goals (TG).

Semantic Abstraction of the System's Response:
{ATTRACTION, INFORM_GENERAL, attraction}system Associated Text Generation Templates: **Option 1**: GENERAL_INFO **Option 2**: PICTURE_INFO Template Selection Rule: IF ({grammar#picture}!=null) THEN select Option 2 ELSE select Option 1
Semantic Abstraction of the System's Response:
{*, BYE, bye}system Associated Text Generation Templates: GOOD_TRIP

5 Evaluation

To evaluate the quality of responses generated by the NLG component, we recruited 15 subjects and asked them to play the role of a tourist in Hong Kong and make related inquiries. The subjects first attend a briefing session where they are presented with the knowledge scope of the system and the supported informational goals (i.e. the 12 task goals in Table 4). The subjects are then instructed to interact with the system textual input and output. The entire interaction is logged and the subjects are subsequently asked to refer to the logged responses (1230 in total) and evaluate each generated response in two ways:

(i) Task Completion Rate – A task is considered complete if the appropriate message exists in the response. For example, if the subject's question is: "迪士尼樂園的門票多少錢?" *(What is the price of a ticket for Disneyland?)* and the system's response is: "迪士尼樂園的票價為成人295，小童210，長者170" *(Ticket prices for Disneyland is 295 for adults, 210 for children and 170 for seniors)* – the response is considered complete. If the subject's question is: "迪士尼樂園的兒童票幾錢？ " *(How much is children's ticket for Disneyland?)* and if the system provides the same answer, the task is also considered complete because the response contains the expected information "小童*210" (210 for children)*. The specificity of the answer is dependent on the current design of the database. It is possible that more specific answers can be generated if the database supports finer granularities in knowledge engineering. Overall 83% of the generated response turns are considered relevant for the task goals based on the user's request turns.

(ii) Grice's Maxims and User Satisfaction – with reference to our previous work [13], we also conducted qualitative evaluation based on Grice's maxims [14] as well as overall user satisfaction. The qualitative evaluation uses a five-point Likert scale (very poor / poor / average / good / very good). Each subject is asked rate the overall quality of the generated responses during his/her interaction with the system, by answering the following questions in a questionnaire:

- **Maxim of Quality**, i.e. system responses should be true with adequate evidence - "*Do you think that the answers are accurate and true?*"
- **Maxim of Quantity**, i.e. system should give sufficient information - "*Do you think that the answers are informative?*"
- **Maxim of Relevance**, i.e. system responses should be relevant to the ongoing conversation - "*Do you think that the answers are relevant to the conversation?*"
- **Maxim of Manner**, i.e. system responses should be brief and clear, with no obscurity or ambiguity - "*Do you think that the answers are clear?*"
- **Overall User Satisfaction** - "*To what extent are you satisfied with the overall performance of the system in responding to your questions?*"

Table 11 shows the average scores and standard derivations (in brackets) of the evaluation results. A *t*-test shows that our results are significantly better than average (Likert score 3) at α=0.06.

Table 11. Evaluation results of our response generation system in terms of Grice's Maxims and user satisfaction.

Quality	Quantity	Relevance	Manner	Satisfaction
4.0 (0.7)	4.1 (0.8)	3.8 (0.7)	3.9 (0.8)	4.0 (0.6)

Analysis of the evaluation logs indicated one common error which accounted to 10% of the incomplete tasks. For example, we found that if the discourse history involved an inquiry with the task goal (TG) of ROUTE, as in the question "怎麼去海洋公園？" *(How to I get to Ocean Park?)* followed by a general question that does not have an obvious TG, e.g. "海洋公園有什麼？" *(What's there in Ocean Park?)* ; then the discourse inheritance mechanism will inherit the TG of ROUTE to the current question, thereby leading to the generation of an incoherent response. Based on the comments offered by the subjects after the evaluation exercise, this kind of error was the main cause of dissatisfaction during the interaction.

6 Conclusions and Future Work

This paper presents a corpus-based approach for cooperative response generation in a spoken dialog system for the Hong Kong tourism domain. A corpus with 3874 requests and responses is collected using Wizard-of-Oz framework. The corpus then undergoes a regularization process that simplifies the interactions to ease subsequent modeling. A semi-automatic process is developed to annotate the each utterance in the dialog turns in terms of their key concepts (KC), task goal (TG) and dialog acts (DA). TG and DA characterize the informational goal and communicative goal of the utterance respectively. The annotation procedure is integrated with a dialog modeling heuristic and a discourse inheritance strategy to generate a semantic abstraction (SA), in the form of $\{TG, DA, KC\}$, for each user request and system response in the dialog. Semantic transitions, i.e. $\{TG, DA, KC\}_{\text{user}} \rightarrow \{TG, DA, KC\}_{\text{system,}}$ may hence be directly derived from the corpus as rules for *response message planning*. Related verbalization methods may also be derived from the corpus and used as templates for *response message realization*. All the rules and templates are stored externally in a human-readable text file which brings the advantage of easy extensibility of the system. Evaluation of this corpus based approach shows that 83% of the generated responses are coherent with the user's request and qualitative rating achieves a score of 4.0 on a five-point Likert scale. Future work will be devoted towards response generation of semantic-dependent expressive markups for text-to-speech synthesis.

Acknowledgments. This work is support by the research fund from the Research Grant Council of Hong Kong (RGC) under grant No. CUHK4237/03E and the joint fund of NSFC (National Natural Science Foundation of China) - RGC under grant No. 60418012 and N-CUHK417/04.

References

1. Hovy, E.: Language Generation. Survey of the State of the Art in Human Language Technology (1996)
2. Bateman, J., Henschel, R.: From Full Generation to "Near-Templates" without losing generality. In: Proc. of the KI'99 Workshop, "May I Speak Freely?" (1999)
3. Heisterkamp, P.: Time to Get Real: Current and Future Requirements for Generation in Speech and Natural Language from an Industrial Perspective. In: Proc. of the KI'99 Workshop, "May I Speak Freely?" (1999)
4. Bateman, J.: KPML Development Environment: Multilingual Linguistic Resource Development and Sentence Generation. German National Center for Information Technology, IPSI (1997)
5. Elhadad, M., Robin, J.: An Overview of SURGE: A Reusable Comprehensive Syntactic Realization Component. Technical Report 96-03, Dept. of Mathematics and Computer Science, Ben Gurion University (1996)
6. Galley, M., Fosler-Lussier, E., Potamianos, A.: Hybrid Natural Language Generation for Spoken Dialogue Systems. In: Proc. of Seventh European Conference on Speech Communication and Technology (Eurospeech '01), Aalborg, Denmark (2001)
7. Young, S.: Talking to Machines (Statistically Speaking). In: Proc. of the International Conference on Spoken Language Processing (2002)
8. Walker, M., Rambow O., Rogati, M.: A Trainable Approach to Sentence Planning for Spoken Dialogue. Computer Speech and Language (2002)
9. Alexandersson, J., Buschbeck-Wolf, Fujinami, M.K., Koch, E.M., Reithinger, B.S.: Acts in VERBMOBIL-2 Second Edition: Verbmobil Report 226, Universitat Hamburg, DFKI Saarbrucken, Universitat Erlangen, TU Berlin
10. Allen, J., Core, M.: Draft of DAMSL: Dialog Act Markup in Several Layers, http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/RevisedManual.html
11. Meng, H., Lam, W., Wai, C.: To Believe is to Understand. In: Proc. of Eurospeech (1999)
12. Chan, S.F., Meng, H.: Interdependencies among Dialog Acts, Task Goals and Discourse Inheritance in Mixed-Initiative Dialogs. In: Proc. of Human Language Technology (2002)
13. Meng, H., Yip, W.L, Mok, O.Y., Chan, S.F.: Natural Language Response Generation in Mixed-Initiative Dialogs using Task Goals and Dialog Acts. In: Proc. of Eurospeech (2003)
14. Frederking, R.: Grices's Maxims: Do the Right Thing. Frederking, R.E. (1996)