

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/073453 A1

(51) International Patent Classification⁷: **G06F 17/20**,
17/28, G10L 5/06, 9/00

(US). **RAMBOW, Owen, Christopher** [US/US]; 833 Garden St., Apt. 1, Hoboken, NJ 07030 (US). **ROGATI, Monica** [RO/US]; 4115 Millington Road, Pittsburgh, PA 15217 (US).

(21) International Application Number: PCT/US02/07235

(22) International Filing Date: 11 March 2002 (11.03.2002)

(74) Agents: **MUDGE, Brian, S.** et al.; Kenyon & Kenyon, 1500 K Street, N.W., Suite 700, Washington, DC 20005 (US).

(25) Filing Language: English

(26) Publication Language: English

(81) Designated States (*national*): CA, US.

(30) Priority Data:
60/275,653 14 March 2001 (14.03.2001) US

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

(71) Applicant (*for all designated States except US*): **AT & T CORP.** [US/US]; 32 Avenue of the Americas, New York, NY 10013-2412 (US).

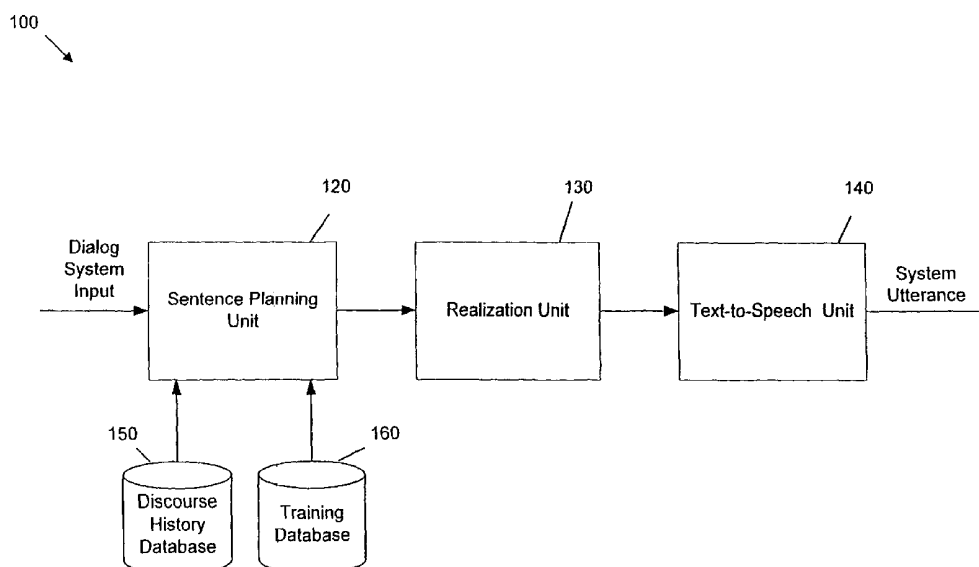
Published:
— with international search report

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **WALKER, Marilyn, A.** [US/US]; 28 Franklin Street, Morristown, NJ 07960

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: A TRAINABLE SENTENCE PLANNING SYSTEM



(57) **Abstract:** The invention relates to a system that interacts with a user in an automated dialog system (100). The system may include a communicative goal generator (210) that generates communicative goals based on a first communication received from the user. The generated communicative goals (210) may be related to information needed to be obtained from the user. The system may further include a sentence planning unit (220) that automatically plans one or more sentences based on the communicative goals generated by the communicative goal generator (210). At least one of the planned sentences may be then output to the user (230).



WO 02/073453 A1

A TRAINABLE SENTENCE PLANNING SYSTEM

CLAIM FOR PRIORITY/CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This non-provisional application claims the benefit of U.S. Provisional Patent Application No. 60/275,653, filed March 14, 2001, which is incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] This invention relates to automated systems for communication recognition and understanding.

BACKGROUND OF THE INVENTION

[0003] The past several years have seen a large increase in commercial spoken dialog systems. These systems typically utilize system-initiative dialog strategies. The system utterances are highly scripted for style and then recorded by voice talent. However several factors argue against the continued use of these simple techniques for producing the system side of the conversation. First, the quality of text-to-speech systems has improved to the point of being a viable alternative to prerecorded prompts.

Second, there is a perceived need for spoken dialog systems to be more flexible and support user initiative. However, this factor also requires greater flexibility for system utterance generation. Finally, there are dialog systems that support complex planning currently under development, and these systems are likely to require more sophisticated system output than current technologies will be able to provide.

SUMMARY OF THE INVENTION

[0004] The invention relates to a system that interacts with a user in an automated dialog system. The system may include a communicative goal generator that generates communicative goals based on a first communication received from the user. The generated communicative goals may be related to information needed to be obtained from the user. The system may further include a sentence planning unit that automatically plans one or more sentences based on the communicative goals generated by the communicative goal generator. At least one of the planned sentences may be then output to the user.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0005] Fig. 1 illustrates an exemplary sentence planning system;
- [0006] Fig. 2 illustrates an exemplary sentence planning unit;
- [0007] Fig. 3 illustrates an exemplary sentence planning system process;
- [0008] Fig. 4 illustrates a list of clause combining operations with examples;
- [0009] Fig. 5 illustrates an alternative zero planning tree;
- [0010] Fig. 6 illustrates an alternative eight sentence planning tree;
- [0011] Fig. 7 illustrates an alternative eight DSYNT structure;
- [0012] Fig. 8 illustrates rules with the largest impact on the final rank booster score;

[0013] Fig. 9 illustrates an exemplary task classification system; and

[0014] Fig. 10 illustrates an exemplary task classification process.

DETAILED DESCRIPTION

[0015] Sentence planning is a set of inter-related but distinct tasks, one of which is sentence scoping. Sentence scoping relates to the choice of syntactic structure for elementary speech acts and the decisions concerning how to combine them into sentences. For example, consider the required capabilities of a sentence planning system for a mixed-initiative spoken dialog system for travel planning in the sample dialog below:

- 1) System: Welcome...What airport would you like to fly out of?
- 2) User: I need to go to Columbus.
- 3) System: Flying to Columbus. What departure airport was that?
- 4) User: From Washington on September the 6th.
- 5) System: What time would you like to travel on September the 6th to Columbus from Washington?

[0016] In utterance 1), the system requests information about the user's departure airport, but in the user's response 2), the user takes the initiative to provide information about a destination. In the system's utterance 3), the system's goal is to implicitly confirm the destination (because of the possibility of error in the speech recognition component), and to also request information (for the second time) concerning the caller's departure airport. In the user's response 4), the caller provides the requested information but also provides the month and day of travel. Given the system's dialog strategy, the communicative goals for the system's utterance 5) are to implicitly confirm all of the information that the user has provided so far, i.e., the departure and destination cities, and the month and day of travel, as well as to request

information about the time of travel. The system's representation of its communicative goals for the system's utterances is illustrated in Table 1 below:

implicit-confirm(orig-city: WASHINGTON)

implicit-confirm(dest-city: COLUMBUS)

implicit-confirm(month: 9)

implicit-confirm(day-number: 6)

request(depart-time: whatever)

Table 1: The Communicative Goals for System Utterance 5, Above.

[0017] An important job for the sentence planning system is to decide among the large number of potential realizations of these communicative goals. Some example alternative realizations are found in Table 2 below:

<u>Alt</u>	<u>Realization</u>	<u>H</u>	<u>MLP</u>
0	What time would you like to travel on September 6 th to Columbus from Washington?	5	.85
5	Leaving on September 6 th . What time would you like to travel from Washington to Columbus?	4.5	.82
8	Leaving in September. Leaving on the 6 th . What time would you, travelling from Washington to Columbus, like to leave?	2	.39

Table 2: Alternative Sentence Plan Realizations for the Communicative Goals for System Utterance 5 in the Sample Dialog, Above

[0018] Fig. 1 illustrates an exemplary sentence planning system 100 which may be used in the above sentence planning scenario as well as in many other various applications, including customer care, service or parts ordering, travel arrangements bookings, location/map information, etc. As shown in the figure, the sentence planning system 100 may include a sentence planning unit 120, a realization unit 130, a text-to-speech unit 140, a discourse history database 150, and a training database 160.

[0019] The sentence planning system 100 receives input recognition and understanding data from a dialog system that is based on input communications from the user. The dialog system may be any system that may be trained to recognize and understand any number of communication symbols, both acoustic and non-acoustic, including grammar fragments, meaningful words, meaningful phrases, meaningful phrase clusters, superwords, morphemes, multimodal signals, etc., using any of the methods known to one skilled in the art including those found in U.S. Patent Nos. 5,675,707, 5,860,063 and 6,044,337, and U.S. Patent Application Nos. 08/943,944, 09/712,192 and 09/712,194, which are hereby incorporated by reference in their entirety. For example, the dialog system may operate using one or more of a variety of recognition and understanding algorithms to determine whether the user's input communications have been recognized and understood prior to inputting data to the sentence planning system 100.

[0020] In the sentence planning system 100, the discourse history database 150 serves as a database for storing each dialog exchange for a particular dialog or set of interactions with a user. The training database 160 stores sentence planning examples collected from interactions with human users and models built based on those examples and positive and negative feedback on the quality of the examples that was provided by human users during the training phase. The training database 150 also stores the sentence planning features identified from the collected dialogs, and the sentence planning rules generated from both the dialogs and the sentence planning features. The sentence planning unit 120 exploits the training database 160 by using the dialog history stored in the discourse history database 150 to predict what sentence plan to generate for the current user interaction.

[0021] While the discourse history database 150 and the training database 160 are shown as separate databases in the exemplary embodiments, the dialog history and training data may be stored in the same database or memory, for example. In any case, any of the databases or memories used by the sentence planning system 100 may be stored external or internal to the system 100.

[0022] Fig. 2 is a more detailed diagram of an exemplary sentence planning unit 120 shown in Fig. 1. The sentence planning unit 120 may include a communicative goal generator 210, a sentence plan generator 220 and a sentence plan ranker 230. The sentence plan generator 220 also receives input from the discourse history database 150 and the sentence plan ranker 230 also receives input from the training database 160.

[0023] The communicative goal generator 210 applies a particular dialog strategy to determine what the communicative goals should be for the system's next dialog turn. Although shown in Fig. 2 as part of the sentence planning unit 120, in another exemplary embodiment (shown by the dotted line), the communicative goal generator 210 may be separate from the sentence planning unit 120 and as such, may be a component of a dialog manager for an automated dialog system, for example (e.g., see Fig. 9). While traditional dialog managers used in conventional spoken dialog systems express communicative goals by looking up string templates that realize these goals and then simply pass the strings to a text-to-speech engine, the communicative goal generator 210 in the present invention generates semantic representations of communicative goals, such as those shown in Table 1.

[0024] These semantic representations are passed to the sentence planning unit 120 that can then use linguistic knowledge and prior training to determine the best realization for these communicative goals given the current discourse context, discourse history, and user. While the communicative goal generator 210 may or may not be physically located in the sentence planning unit 120, or even be a part of the sentence planning system 100, within the spirit and scope of the invention, for ease of discussion, the communicative goal generator 210 will be discussed as being part of the sentence planning unit 120.

[0025] In order to train the sentence planning system 100, the sentence planning process may include two distinct phases performed by the sentence plan generator 220 and the sentence plan ranker 230, respectively. In the first phase, the sentence plan generator 210 generates a potentially large sample of possible sentence plans for a

given set of communicative goals generated by the communicative goal generator 210. In the second phase, the sentence-plan-ranker 220 ranks the sample sentence plans and then selects the top ranked plan to input to the realization unit 130. In ranking the generated sentence plans, the sentence plan ranker 230 may use rules automatically learned from training data stored in the training database 160, using techniques similar to those well-known to one of ordinary skill in the art.

[0026] In order to train the sentence planning system 100, neither hand-crafted rules nor the existence of a corpus in the domain of the sentence planning system 100 are necessarily needed. The trained sentence plan ranker 230 may learn to select a sentence plan whose rating on average is only 5% worse than the top human-ranked sentence plan. To further illustrate this, the sentence planning process, as well as the detailed descriptions of the sentence plan generator 220 and the sentence plan ranker 230, is set forth below.

[0027] Fig. 3 illustrates an exemplary sentence planning process using the sentence planning system 100. The process begins at step 3005 and proceeds to step 3010 where the communicative goal generator 210 receives recognition and understanding data from a dialog system and calculates the communicative goals of the particular transaction with the user. In step 3020, the communicative goal generator 210 transfers the calculated communicative goals along with the recognized/understood symbols to the sentence planning generator 220. The sentence plan generator 220 uses inputs from the discourse history database 150 to generate a plurality of sentence plans. Then, in step 3030, the generated sentence plans are ranked by the sentence plan ranker 230 using a set of rules stored in the training database 160.

[0028] The process proceeds to step 3040 where the sentence plan ranker 230 selects the highest ranked sentence plan. In step 3050, the selected sentence plan is input to the realization unit 130, which may be either a rule-based or stochastic surface realizer, for example. In the realization unit 130, linguistic rules and/or linguistic knowledge, derived from being trained using an appropriate dialog corpus, are applied to generate the surface string representation. Specifically, the types of linguistic rules or

knowledge that the realization unit 130 may apply may concern the appropriate irregular verb forms, subject-verb agreement, inflecting words, word order, and the application of function words. For example, in English, the indirect object of the verb "give" is matched with the function word "to" as in the sentence "Matthew GAVE the book TO Megan". Note that for ease of discussion, "linguistic rules" as described herein will be intended to encompass either or both "linguistic rules" and/or "linguistic knowledge".

[0029] Then, in step 3060, the realized sentence plan is converted from text to speech by the text-to-speech unit 140 and is output to the user in step 3070. The text-to-speech unit 140 may be a text-to-speech engine known to those of skill in the art, such as that embodied in the AT&T NextGen TTS system, and possibly trained with lexical items specific to the domain of the sentence planning system 100. The device that outputs the converted sentence may be any device capable of producing verbal and/or non-verbal communications, such as a speaker, transducer, TV screen, CRT, or any other output device known to those of ordinary skill in the art. If the output includes speech, the automated speech may be produced by a voice synthesizer, voice recordings, or any other method or device capable of automatically producing audible sound known to those of ordinary skill in the art. The process then goes to step 3080 and ends.

[0030] In general, the role of the sentence planning system 100 is to choose abstract lexico-structural realizations for a set of communicative goals generated by the communicative goal generator 210. In contrast to conventional dialog systems that simply output completely formed utterances, the output of the above-described text-to-speech unit 140 provides the input back to the sentence planning system 100 in the form of a single spoken dialog text plan for each interaction between the system and the user.

[0031] In this process, each sentence plan generated by the sentence plan generator 220 is an unordered set of elementary speech acts encoding all of the communicative goals determined by the communicative goal generator 210 for the current user interaction. As illustrated above in Table 1, each elementary speech act is

represented as a type (request, implicit confirm, explicit confirm), with type-specific parameters. The sentence planning system 100 must decide among alternative realizations of this communicative goal. Some alternative realizations are shown in Table 2, above.

[0032] As discussed above, the sentence planning task is divided by the sentence planning unit 120 into two phases. In the first phase, the sentence plan generator 220 generates 12-20 possible sentence plans, for example, for a given input communicative goal. To accomplish this, the sentence plan generator 220 assigns each speech act a canonical lexico-structural representation called a "Deep Syntactic Structure" (DSyntS). Essentially, the sentence plan is a tree that records how these elementary DSyntSs are combined into larger DSyntSs. From a sentence plan, the list of DSyntSs, each corresponding to exactly one sentence of the target communicative goal, can be read off. In the second phase, the sentence plan ranker 230 ranks sentence plans generated by the sentence plan generator 220, and then selects the top-ranked output which is then input into the realization unit 130.

[0033] In examining each of these phases, the sentence plan generator 220 performs a set of clause-combining operations that incrementally transform a list of elementary predicate-argument representations (the DSyntSs corresponding to elementary speech acts, in this case) into a list of lexico-structural representations of single sentences. As shown in Fig. 4, the sentence plan generator 220 performs this task by combining the elementary predicate-argument representations using the following combining operations:

- MERGE. Two identical main matrix verbs can be identified if they have the same arguments; the adjuncts are combined.
- MERGE-GENERAL. Same as MERGE, except that one of the two verbs may be embedded.
- SOFT-MERGE. Same as MERGE, except that the verbs need only to be in a relation of synonymy or hyperonymy (rather than being identical).
- SOFT-MERGE-GENERAL. Same as MERGE-GENERAL, except that the verbs need only to be in a relation of synonymy or hyperonymy.

- CONJUNCTION. This is standard conjunction with conjunction reduction.
- RELATIVE-CLAUSE. This includes participial adjuncts to nouns.
- ADJECTIVE. This transforms a predicative use of an adjective into an adnominal construction.
- PERIOD. Joins two complete clauses with a period.

[0034] The output of the sentence plan generator 220 is a sentence plan tree (or sp-tree for short), which is a binary tree with leaves labeled by all the elementary speech acts from the input communicative goals, and with its interior nodes labeled with clause-combining operations. Each node is also associated with a DSyntS: the leaves which correspond to elementary speech acts from the input communicative goals are linked to a canonical DSyntS for that speech act by lookup in a hand-crafted dictionary, for example. The interior nodes are associated with DSyntSs by executing their clause-combining operation on their two daughter nodes. For example, a *PERIOD* node results in a DSyntS headed by a period and whose daughters are the two daughter DSyntSs. As a result, the DSyntS for the entire user interaction is associated with the root node. This DSyntS can be sent to the realization unit 130, which outputs a single sentence or several sentences if the DSyntS contains period nodes.

[0035] The complexity of conventional sentence planning systems arises from the attempt to encode constraints on the application and ordering of system operations in order to generate a single high-quality sentence plan. However, in the sentence planning system 100 process of the invention there is not a need to encode such constraints. Instead, the sentence plan generator 220 generates a random sample of possible sentence plans for each communicative goal generated by the communicative goal generator 210. This may be accomplished by randomly selecting among the operations according to a probability distribution. If a clause combination fails, the sentence plan generator 220 discards that sp-tree. For example, if a relative clause of a structure which already contains a period is created, it will be discarded.

[0036] Table 2 above shows some of the realizations of alternative sentence plans generated by the sentence plan generator 220 for utterance systems in the sample dialog above. Sp-trees for alternatives 0, 5 and 8 are shown in Figs. 5 and 6.

For example, consider the sp-tree in Fig. 6. Node soft-merge-general merges an implicit-confirmation of the destination city and the origin city. The row labeled *SOFT-MERGE* in Fig. 4 shows the result of applying the soft-merge operation when Args 1 and 2 are implicit confirmations of the origin and destination cities. Fig. 7 illustrates the relationship between the sp-tree and the DSynt structure for alternative 8 from Fig. 6. The labels and arrows show the DSynt structure associated with each node in the sp-tree. The Fig. 7 diagram also shows how structures are composed into larger structures by the clause-combining operations.

[0037] The sentence plan ranker 230 takes as input a set of sentence plans generated by the sentence plan generator 220 and ranks them. As discussed above, in order to train the sentence plan ranker 230, a machine learning program may be applied to learn a set of rules for ranking sentence plans from the labeled set of sentence-plan training examples stored in the training database 160.

[0038] Examples of boosting algorithms that may be used by the sentence plan ranker 230 for ranking the generated sentence plans are described in detail below. Each example x is represented by a set of m indicator functions $h_s(x)$ for $1 \leq s \leq m$. The indicator functions are calculated by thresholding the feature values (counts) described below. For example, one such indicator function might be:

$$h_{100}(x) = \begin{cases} 1 & \text{if } \text{LEAF-IMPLICIT-CONFIRM}(x) \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

[0039] So $h_{100} = 1$ if the number of leaf implicit confirm nodes in $x \geq 1$. A single parameter a_s is associated with each indicator function, and the "ranking score" for an example x is then calculated as:

$$F(x) = \sum_s a_s h_s(x)$$

[0040] The sentence plan ranker 230 uses this score to rank competing realizations of the same text plan in order of plausibility. The training examples are used to set the parameter values a_s . In this case, the human judgments are converted

into a training set of *ordered pairs* of examples x, y , where x and y are candidates for the same sentence, and x is strictly preferred to y . More formally, the training set τ is:

$$\tau = \{(x, y) \mid x, y \text{ are realizations for the same text plan,}$$

$x \text{ is preferred to } y \text{ by human judgments}\}$

Thus, each text plan with 20 candidates could contribute up to $(20 * 19)/2 = 190$ such pairs. In practice, however, fewer pairs could be contributed due to different candidates getting tied scores from the annotators.

[0041] Training is then described as a process of setting the parameters a_s to minimize the following loss function:

$$Loss = \sum_{(x,y) \in \tau} e^{(F(x)-F(y))}$$

[0042] It can be seen that as this loss function is minimized, the values for $(F(x) - F(y))$ where x is preferred to y will be pushed to be positive, so that the number of ranking errors (cases where ranking scores disagree with human judgments) will tend to be reduced. Initially all parameter values are set to zero. The optimization method then picks a single parameter at a time, preferably the parameter that will make most impact on the loss function, and updates the parameter value to minimize the loss. The result is that substantial progress is typically made in minimizing the error rate, with relatively few non-zero parameter values. Consequently, under certain conditions, the combination of minimizing the loss function while using relatively few parameters leads to good generalization on test data examples. Empirical results for boosting have shown that in practice the method is highly effective.

[0043] Fig. 8 shows some of the rules that were learned on the training data that were then applied to the alternative sentence plans in each test set of each fold in order to rank them. Only a subset of the rules that had the largest impact on the score of each sp-tree is listed. Some particular rule examples are discussed here to help in

understanding how the sentence plan ranker 230 operates. However, different thresholds and feature values may be used within the spirit and scope of the invention.

[0044] Rule (1) in Fig. 8 states that an implicit confirmation as the first leaf of the sp-tree leads to a large (.94) increase in the score. Thus, all three of the alternative sp-trees accrue this ranking increase. Rules (2) and (5) state that the occurrence of 2 or more PRONOUN nodes in the DsyntS reduces the ranking by 0.85, and that 3 or more PRONOUN nodes reduces the ranking by an additional 0.34. Alternative 8 is above the threshold for both of these rules; alternative 5 is above the threshold for Rule (2) and alternative 0 is never above the threshold. Rule (6) on the other hand increases only the scores of alternatives 0 and 5 by 0.33 since alternative 8 is below threshold for that feature.

[0045] Although multiple instantiations of features are provided, some of which included parameters or lexical items that might identify particular discourse contexts, most of the learned rules utilize general properties of the sp-tree and the DSyntS. This is partly due to the fact that features that appeared less than 10 times in the training data were eliminated.

[0046] Fig. 9 shows an exemplary task classification system 900 that includes the sentence planning system 100. The task classification system 900 may include a recognizer 920, an NLU unit 930, a dialog manager/task classification processor 940, a sentence planning unit 120, a realization unit 130, a text-to-speech unit 140, a discourse history database 150, and a training database 160. The functions and descriptions of the sentence planning unit 120, the realization unit 130, the text-to-speech unit 140, the discourse history database 150, and the training database 160 are set forth above and will not be repeated here.

[0047] The sentence planning unit 120 receives recognition data from the recognizer 920 and understanding data from the NLU unit 930 that are based on input communications from the user. The recognizer 920 and the NLU unit 930 are shown as separate units for clarification purposes. However, the functions of the recognizer 920

and the NLU unit 930 may be performed by a single unit within the spirit and scope of this invention.

[0048] Note that the recognizer 920 may be trained to recognize any number of communication symbols, both acoustic and non-acoustic, including grammar fragments, meaningful words, meaningful phrases, meaningful phrase clusters, superwords, morphemes, multimodal signals, etc., using any of the methods known to one skilled in the art including those found in U.S. Patent Nos. 5,675,707, 5,860,063 and 6,044,337, and U.S. Patent Application Nos. 08/943,944, 09/712,192 and 09/712,194, as discussed above.

[0049] The recognizer 920 and the NLU unit 930 may operate using one or more of a variety of recognition and understanding algorithms. For example, the recognizer 920 and the NLU unit 930 may use confidence functions to determine whether the user's input communications have been recognized and understood. The recognition and understanding data from the user's input communication may be used by the NLU unit 930 to calculate a probability that the language is understood clearly and this may be used in conjunction with other mechanisms like recognition confidence scores to decide whether and/or how to further process the user's communication.

[0050] The dialog manager/task classification processor 940 may be used to solicit clarifying information from the user in order to clear up any system misunderstanding. As a result, if the user's input communication can be satisfactorily recognized by the recognizer 920, understood by the NLU unit 930, and no further information from the user is needed, the dialog manager/task classification processor 940 routes and/or processes the user's input communication, which may include a request, comment, etc. However, if the NLU unit 930 recognizes errors in the understanding of the user's input communication such that it cannot be satisfactorily recognized and understood, the dialog manager/task classification processor 940 may conduct dialog with the user for clarification and confirmation purposes.

[0051] The dialog manager/task classification processor 940 also may determine whether all of the communicative goals have been satisfied. Therefore, once the

system has collected all of the necessary information from the user, the dialog manager/task classification processor 940 may classify and route any request or task received from the user so that it may be completed or processed by another system, unit, etc. Alternatively, the dialog manager/task classification processor 940 may process, classify or complete the task itself.

[0052] Note that while Fig. 9 shows the dialog manager/task classification processor 940 as a single unit, the functions of the dialog manager portion and the task classification processor portion may be performed by a separate dialog manager and a separate task classification processor, respectively.

[0053] As noted above, the dialog manager/task classification processor 940 may include, or perform the functions of, the communicative goal generator 210. In this regard, the dialog manager/task classification processor 940 would determine the communicative goals based on the recognized symbols and understanding data and route the communicative goals to the sentence plan generator 220 of the sentence planning unit 120.

[0054] Fig. 10 illustrates an exemplary sentence planning process in the task classification system 900. The process begins at step 10005 and proceeds to step 10010 where the recognizer 920 receives an input communication from the user recognizes symbols from the user's input communications using a recognition algorithm known to those of skill in the art. Then, in step 10015, recognized symbols are input to the NLU unit 930 where an understanding algorithm may be applied to the recognized symbols as known to those of skill in the art.

[0055] In step 10020, the NLU unit 930 determines whether the symbols can be understood. If the symbols cannot be understood, the process proceeds to step 10025 where dialog manager/task classification processor 940 conducts dialog with the user to clarify the system's understanding. The process reverts back to step 10010 and the system waits to receive additional input from the user.

[0056] However, if the symbols can be understood in step 10020, the process proceeds to step 10030 where the dialog manager/task classification processor 940 (or the communicative goal generator 210) determines whether the communicative goals in the user transaction have been met. If so, in step 10070, the dialog manager/task classification processor 940 routes the tasks from user's request to another unit for task completion, or processes the user's communication or request, itself. The process then goes to step 10070 and ends.

[0057] If the dialog manager/task classification processor 940 determines whether the communicative goals in the user transaction have not been met, the process proceeds to step 10035 where the communicative goal generator 210 (or the dialog manager/task classification processor 940) calculates the communicative goals of the particular transaction with the user using the recognition and understanding data. In step 10040, the communicative goal generator 210 transfers the calculated communicative goals along with the recognition and understanding data to the sentence planning unit 120. In the sentence planning unit 120, sentence plans are generated by the sentence plan generator 220 using input from the discourse history database 150. Then, in step 10045, the generated sentence plans are ranked by the sentence planning ranker 230.

[0058] The process proceeds to step 10050 where the sentence plan ranker 230 selects the highest ranked sentence plan. In step 10055, the selected sentence plan is input to the realization unit 130 where linguistic rules are applied. Then, in step 10060, the realized sentence plan is converted from text to speech by the text-to-speech unit 140 and is output to the user in step 10065. The process then goes to step 10070 and ends.

[0059] In the discussion herein, the terms "natural language understanding" and "sentence planning" are used to describe the understanding of a user's communication and the automated formulation of a system response, respectively. As such, this invention is directed toward the use of any form of communications received or transmitted over the networks which may be expressed verbally, nonverbally,

multimodally, etc. Examples of nonverbal communications include the use of gestures, body movements, head movements, non-responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes, pointers, stylus, cable set-top box entries, graphical user interface entries, touchscreen entries, etc. Multimodal communications involve communications on a plurality of channels, such as aural, visual, etc. However, for ease of discussion, examples and discussions of the method and system of the invention are discussed above in relation to, but not limited to, verbal systems.

[0060] Note that while the above examples illustrate the invention in a travel service system, this invention may be applied to any single mode, or multimodal, dialog system, or any other automated dialog system that interacts with a user. Furthermore, the invention may apply to any automated recognition and understanding system that receives communications from external sources, such as users, customers, service providers, associates, etc. Consequently, the method may operate in conjunction with one or more communication networks, including a telephone network, the Internet, an intranet, Cable TV network, a local area network (LAN), a wireless communication network, etc.

[0061] In addition, while the examples above concern travel service systems, the sentence planning system 100 of the invention may be used in a wide variety of systems or purposes known to those of skill in the art, including parts ordering systems, customer care systems, reservation systems (including dining, car, train, airline, bus, lodging, travel, touring, etc.), navigation systems, information collecting systems, information retrieval systems, etc., within the spirit and scope of the invention.

[0062] As shown in Figs. 1, 2, and 9, the method of this invention may be implemented using a programmed processor. However, the method can also be implemented on a general-purpose or a special purpose computer, a programmed microprocessor or microcontroller, peripheral integrated circuit elements, an application-specific integrated circuit (ASIC) or other integrated circuits, hardware/electronic logic circuits, such as a discrete element circuit, a programmable logic device, such as a PLD, PLA, FPGA, or PAL, or the like. In general, any device on which the finite state

machine capable of implementing the flowcharts shown in Figs. 3 and 10 can be used to implement the functions of this invention.

[0063] While the invention has been described with reference to the above embodiments, it is to be understood that these embodiments are purely exemplary in nature. Thus, the invention is not restricted to the particular forms shown in the foregoing embodiments. Various modifications and alterations can be made thereto without departing from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

- 1 1. A system that interacts with a user using an automated dialog
2 system, comprising:
3 a communicative goal generator that generates communicative goals
4 based on a first communication received from the user, the generated
5 communicative goals being related to information needed to be obtained from the
6 user; and
7 a sentence planning unit that automatically plans one or more sentences
8 based on the communicative goals generated by the communicative goal
9 generator, wherein at least one of the one or more planned sentences are output
10 to the user.
- 1 2. The system of claim 1, further comprising:
2 a sentence plan generator that generates a plurality of sentence plans
3 based on the generated communicative goals;
4 a sentence plan ranker that ranks the plurality of sentence plans
5 generated by the sentence planning generator and selects the highest ranked
6 sentence plan to be output to the user.
- 1 3. The system of claim 2, further comprising:
2 a realization unit that applies a set of linguistic rules to the sentence plan
3 selected by the sentence plan ranker.
- 1 4. The system of claim 2, further comprising:
2 a training database that includes a set of learned rules, wherein the
3 sentence plan ranker ranks the sentence plan generated by the sentence plan
4 generator using the set of learned rules.
- 1 5. The system of claim 2, further comprising:
2 a discourse history database that includes interaction information related
3 to a set of interactions between the user and the automated dialog system,
4 wherein the sentence plan generator generates a plurality of sentence plans
5 using the interaction information.

1 6. The system of claim 1, wherein the user's first communication
2 includes nonverbal communications.

1 7. The system of claim 6, wherein the nonverbal communications
2 include at least one of gestures, body movements, head movements, non-
3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,
4 pointers, stylus, cable set-top box entries, graphical user interface entries, and
5 touchscreen entries.

1 8. The system of claim 1, wherein the communicative goal generator
2 generates communicative goals using recognition and understanding data from
3 the user's communication.

1 9. The system of claim 1, wherein the communicative goals generated
2 by the communicative goal generator include confirming information previously
3 obtained from the user.

1 10. The system of claim 1, further comprising:
2 a text-to-speech converter that converts at least one of the one or more
3 sentence plans from text to speech.

1 11. The system of claim 1, wherein the system is used in one of a
2 customer care system, a reservation system, parts ordering system, navigation
3 system, information gathering system, and information retrieval system.

1 12. An automated sentence planning system that automatically plans
2 sentences based on communicative goals input from an automated dialog
3 system, the communicative goals being related to information needed to be
4 obtained from a user, comprising:

5 a sentence plan generator that generates a plurality of sentence plans
6 based on the communicative goals;

7 a sentence plan ranker that ranks the sentence plans generated by the
8 sentence plan generator and selects the highest ranked sentence plan; and

9 a realization unit that realizes the selected sentence plan, wherein the
10 realized sentence plan is output to the user.

1 13. The system of claim 12, wherein the realization unit applies a set of
2 linguistic rules to the selected sentence plan.

1 14. The system of claim 12, further comprising:

2 a training database that includes a set of learned rules, wherein sentence
3 plan ranker ranks the generated sentence plans using the set of learned rules.

1 15. The system of claim 12, further comprising:

2 a discourse history database that includes interaction information related
3 to a set of interactions between the user and the automated dialog system,
4 wherein the sentence plan generator generates a plurality of sentence plans
5 using the interaction information.

1 16. The system of claim 12, wherein the interactions between the user
2 and the automated dialog system include nonverbal communications.

1 17. The system of claim 16, wherein the nonverbal communications
2 include at least one of gestures, body movements, head movements, non-
3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,
4 pointers, stylus, cable set-top box entries, graphical user interface entries, and
5 touchscreen entries.

1 18. The system of claim 12, wherein the communicative goals are
2 generated using recognition and understanding data from a communication
3 received by the automated dialog system from the user.

1 19. The system of claim 12, wherein the communicative goals include
2 confirming information previously obtained from the user.

1 20. The system of claim 12, further comprising:
2 a text-to-speech converter that converts the realized sentence plan from
3 text to speech.

1 21. The system of claim 12, wherein the system is used in one of a
2 customer care system, a reservation system, parts ordering system, navigation
3 system, information gathering system, and information retrieval system.

1 22. An automated sentence planning system coupled to an automated
2 dialog system that automatically plans sentences based on communicative goals
3 related to information needed to be obtained from a user, comprising:

4 a sentence plan generator that generates a plurality of sentence plans
5 based on the communicative goals and information related to a set of interactions
6 between the user and the automated dialog system;

7 a sentence plan ranker that ranks the sentence plans generated by the
8 sentence plan generator using a set of learned rules and selects the highest
9 ranked sentence plan;

10 a realization unit that realizes the selected sentence plan by applying a set
11 of linguistic rules; and

12 a text-to-speech converter that converts the realized sentence plan from
13 text to speech, wherein the converted sentence plan is output to the user.

1 23. The system of claim 22 wherein the interactions between the user
2 and the automated dialog system include nonverbal communications.

1 24. The system of claim 23, wherein the nonverbal communications
2 include at least one of gestures, body movements, head movements, non-
3 responses, text, keyboard entries, keypad entries, mouse clicks, DTMF codes,
4 pointers, stylus, cable set-top box entries, graphical user interface entries, and
5 touchscreen entries.

1 25. The system of claim 22, further comprising:
2 a communicative goal generator that generates the communicative goals
3 using recognition and understanding data from a communication received by the
4 automated dialog system from the user.

1 26. The system of claim 22, wherein the communicative goals include
2 confirming information previously obtained from the user.

1 27. The system of claim 22, wherein the system is used in one of a
2 customer care system, a reservation system, parts ordering system, navigation
3 system, information gathering system, and information retrieval system.

1/10

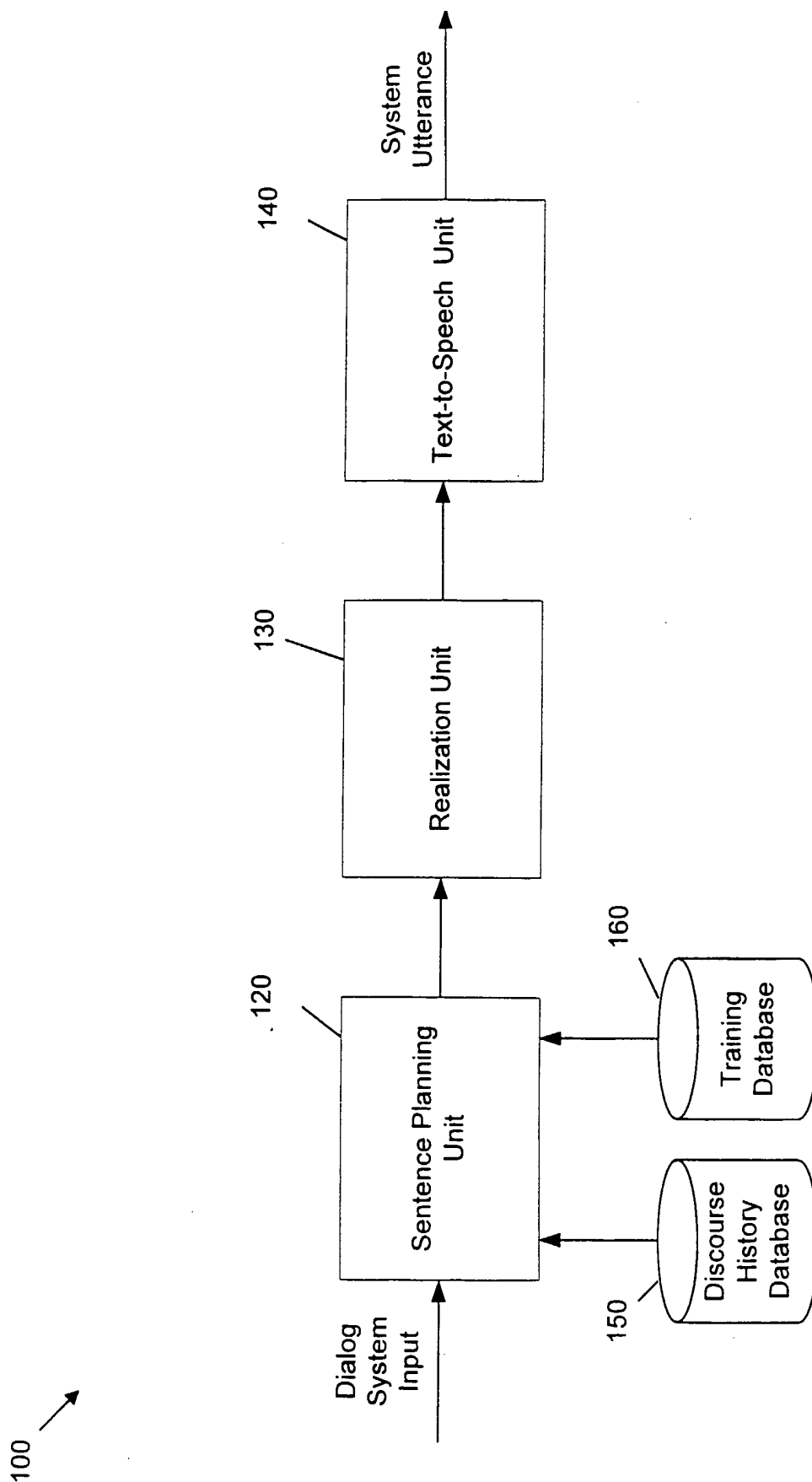


Fig. 1

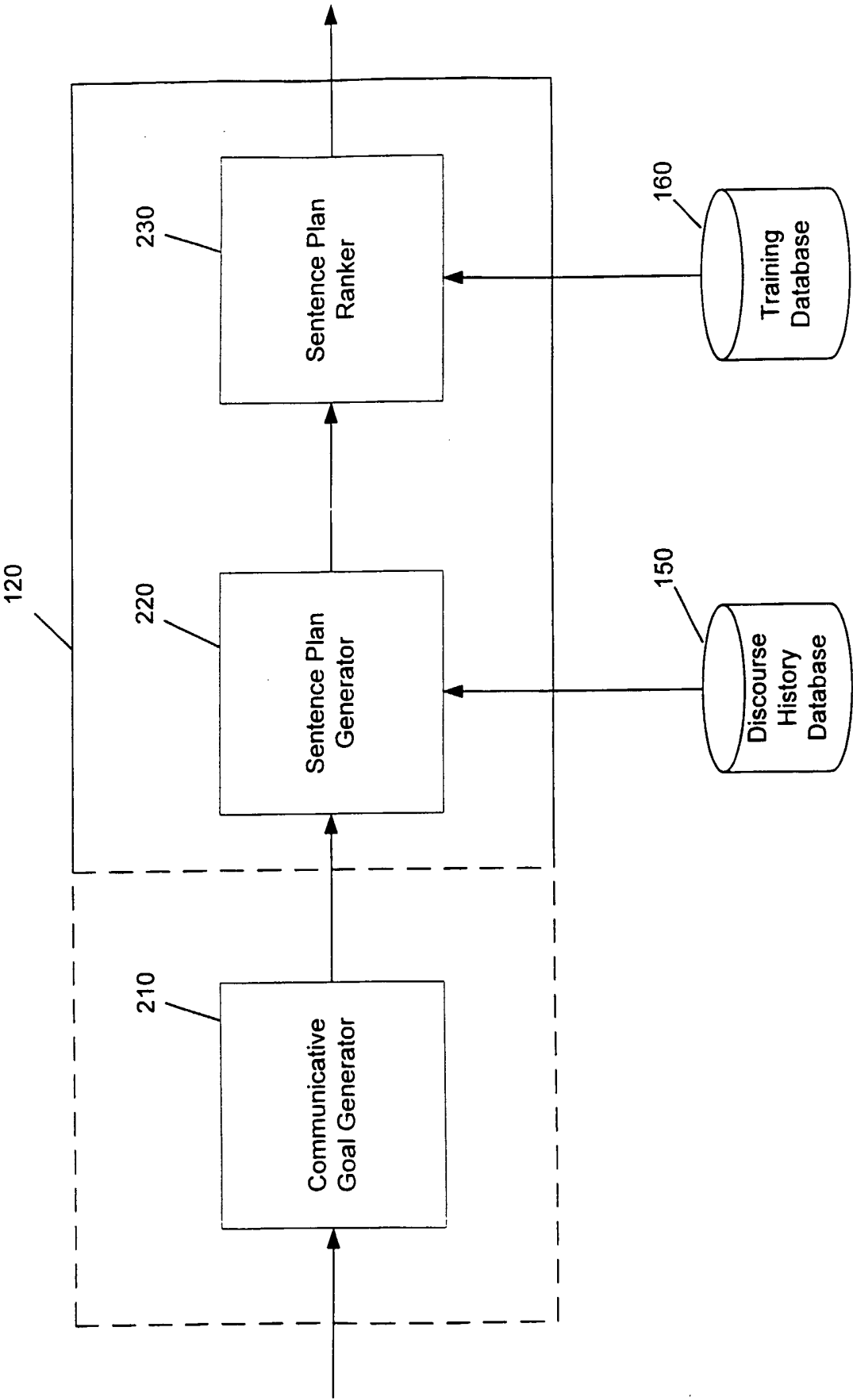
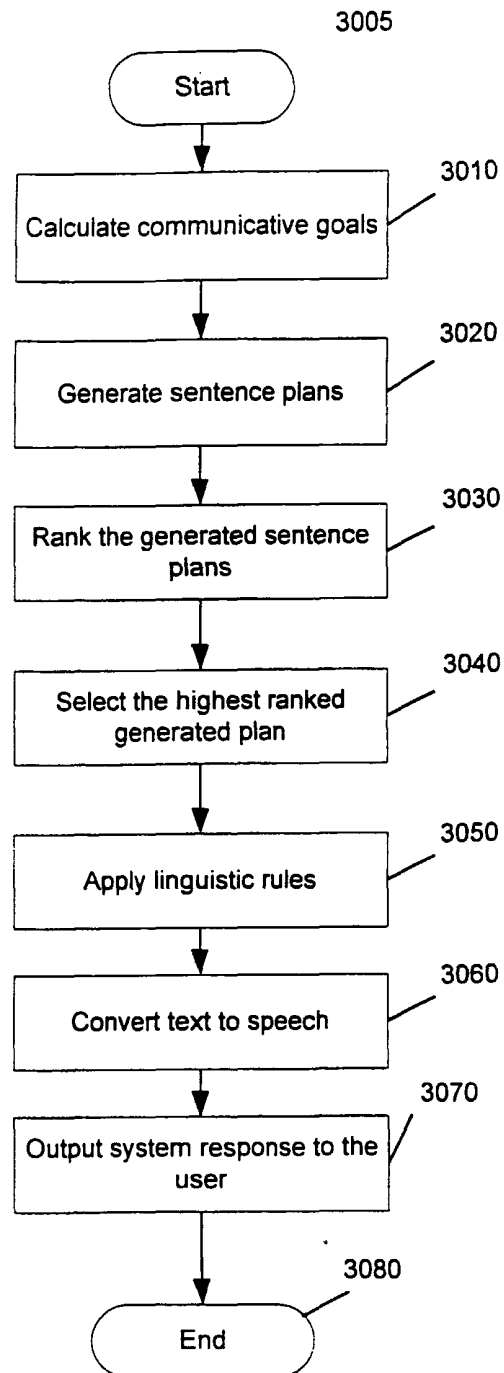


Fig. 2

3/10**Fig. 3**

4/10

Rule	Arg 1	Arg 2	Result
MERGE	You are leaving from Newark.	You are leaving at 5	You are leaving at 5 from Newark
MERGE-GENERAL	What time would you like to leave?	You are leaving from Newark.	What time would you like to leave from Newark?
SOFT-MERGE	You are leaving from Newark	You are going to Dallas	You are traveling from Newark to Dallas
SOFT-MERGE-GENERAL	What time would you like to leave?	You are going to Dallas.	What time would you like to fly to Dallas?
CONJUNCTION	You are leaving from Newark.	You are going to Dallas.	You are leaving from Newark and you are going to Dallas.
RELATIVE-CLAUSE	Your flight leaves at 5.	Your flight arrives at 9.	Your flight, which leaves at 5, arrives at 9.
ADJECTIVE	Your flight leaves at 5.	Your flight is nonstop.	Your nonstop flight leaves at 5.
PERIOD	You are leaving from Newark.	You are going to Dallas.	You are leaving from Newark. You are going to Dallas

Fig. 4

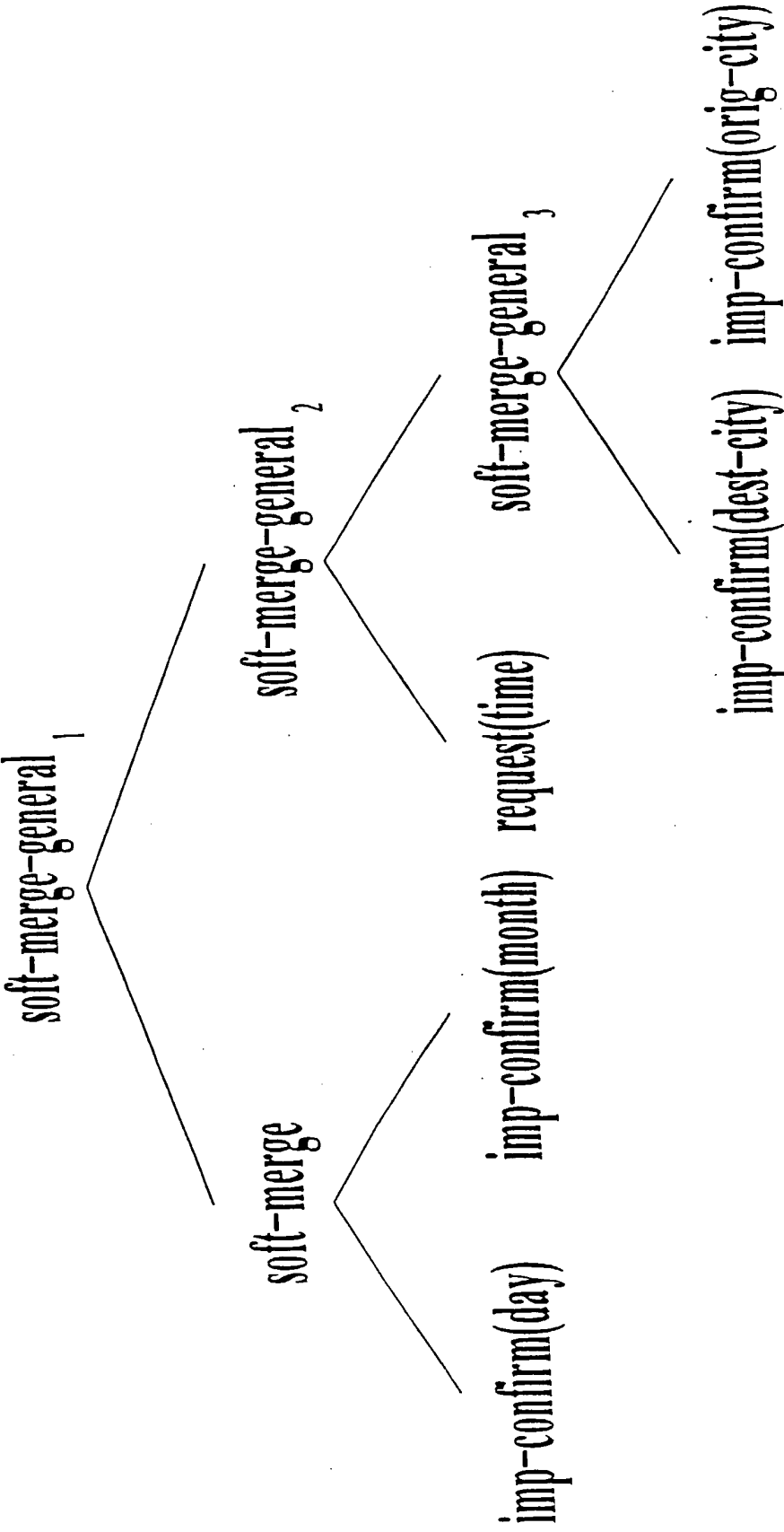


Fig. 5

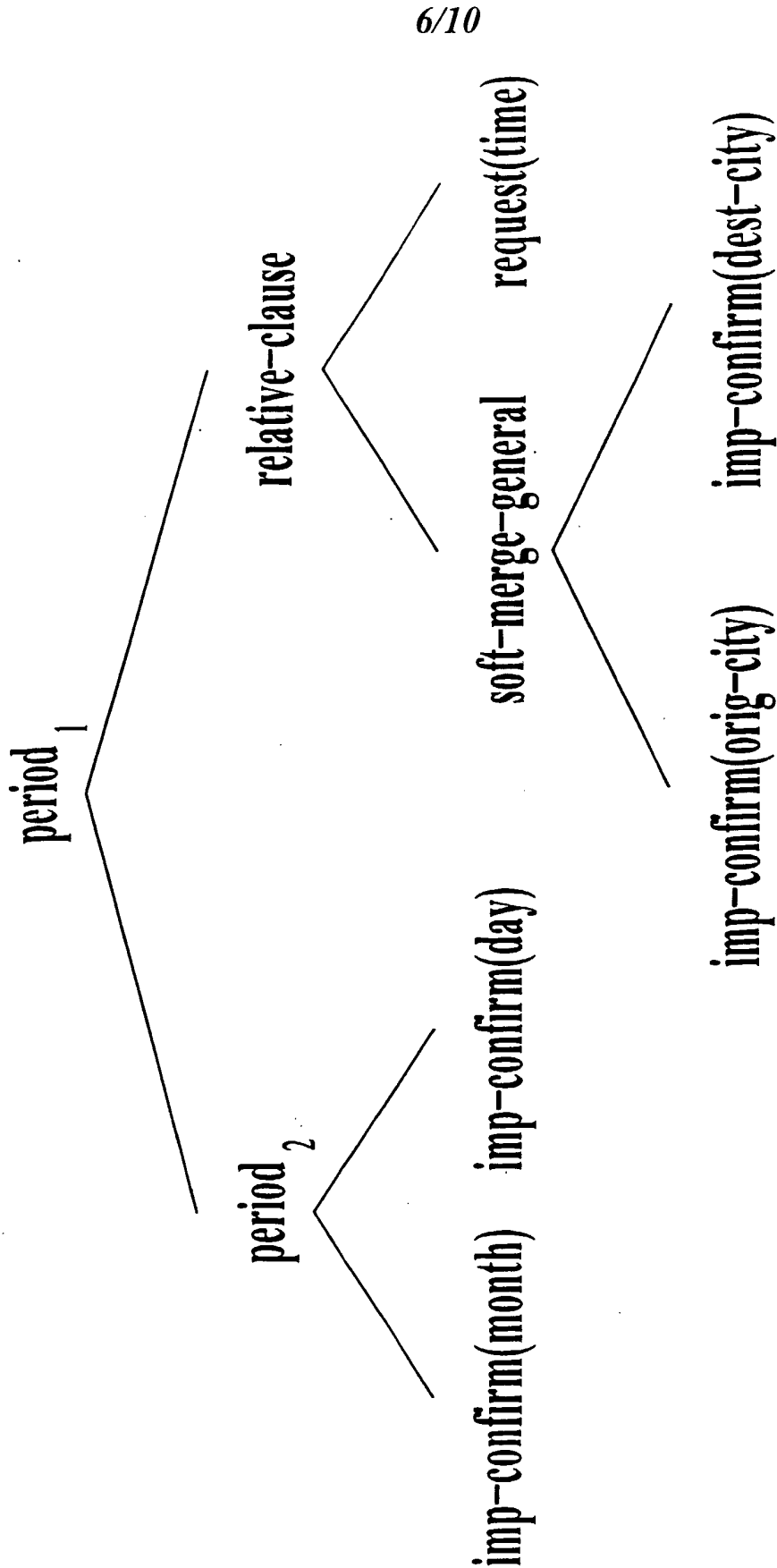


Fig. 6

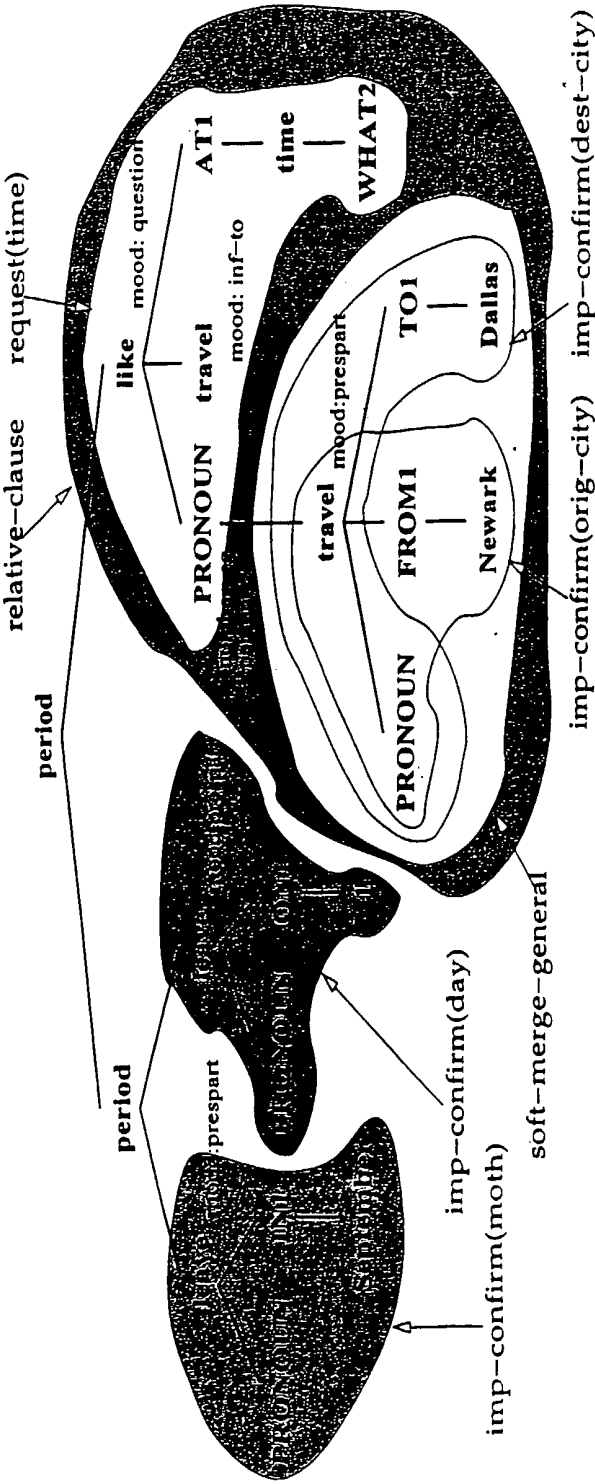


Fig. 7

8/10

N	Condition	A0	A5	A8	α_s
1	LEAF-IMPLICIT-CONFIRM ≥ 1	1	1	1	0.94
2	DSYNT-TRAVERSAL-PRONOUN ≥ 2	1	2	4	-0.85
3	LEAF-IMPLICIT-CONFIRM*IMPLICIT-CONFIRM*REQUEST*IMPLICIT-CONFIRM ≥ 1	1	0	0	-0.52
4	DSYNT-TRAVERSAL-INIT ≥ 1	0	0	1	-0.52
5	DSYNT-TRAVERSAL-PRONOUN ≥ 3	1	2	4	-0.34
6	SP-ANCESTOR*IMPLICIT-CONFIRM-ORIG-CITY*SOFT-MERGE-GENERAL*SOFT-MERGE-GENERAL ≥ 1.0	1	1	0	0.33
7	SP-ANCESTOR-SOFT-MERGE-GENERAL*PERIOD ≥ 1	0	1	0	0.21
8	DSYNT-ANCESTOR-INIT*LEAVE ≥ 1	0	0	1	-0.16
9	SP-TRAVERSAL-IMPLICIT-CONFIRM-DAY-NUMBER ≥ 1	1	1	1	-0.13
10	SP-TRAVERSAL-SOFT-MERGE*IMPLICIT-CONFIRM*IMPLICIT-CONFIRM ≥ 1	1	0	0	0.11
11	REL-CLAUSE-AVG ≥ 2	0	0	3	-0.12
12	PERIOD-AVG ≥ 3	0	5	3.5	0.12
13	DSYNT-ANCESTOR-TRAVEL*LIKE ≥ 1	1	0	0	0.10
14	DSYNT-SISTERS-PRONOUN-ON1 ≥ 1	0	1	1	-0.10
15	LEAF-IMPLICIT-CONFIRM*IMPLICIT-CONFIRM*REQUEST ≥ 1	1	0	0	-0.10
16	REL-CLAUSE-MIN ≥ 2	0	0	3	-0.09
17	SP-SISTERS-IMPLICIT-CONFIRM*IMPLICIT-CONFIRM- ≥ 1	2	2	2	0.09
18	REL-CLAUSE-MAX ≥ 2	0	0	3	-0.07
19	SP-ANCESTOR-IMPLICIT-CONFIRM*SOFT-MERGE*SOFT-MERGE-GENERAL ≥ 1	1	0	0	0.06

Fig. 8

9/10

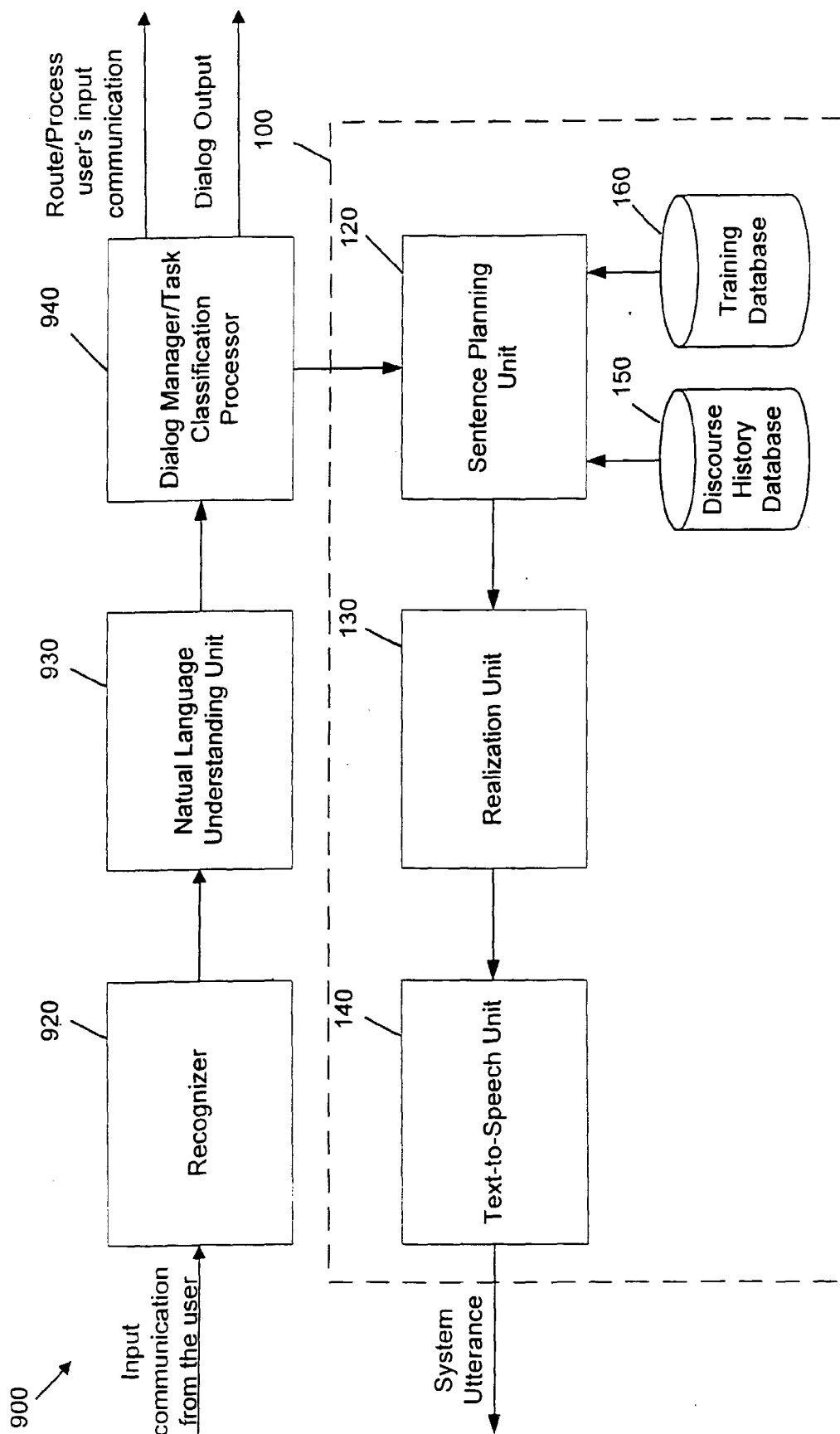


Fig. 9

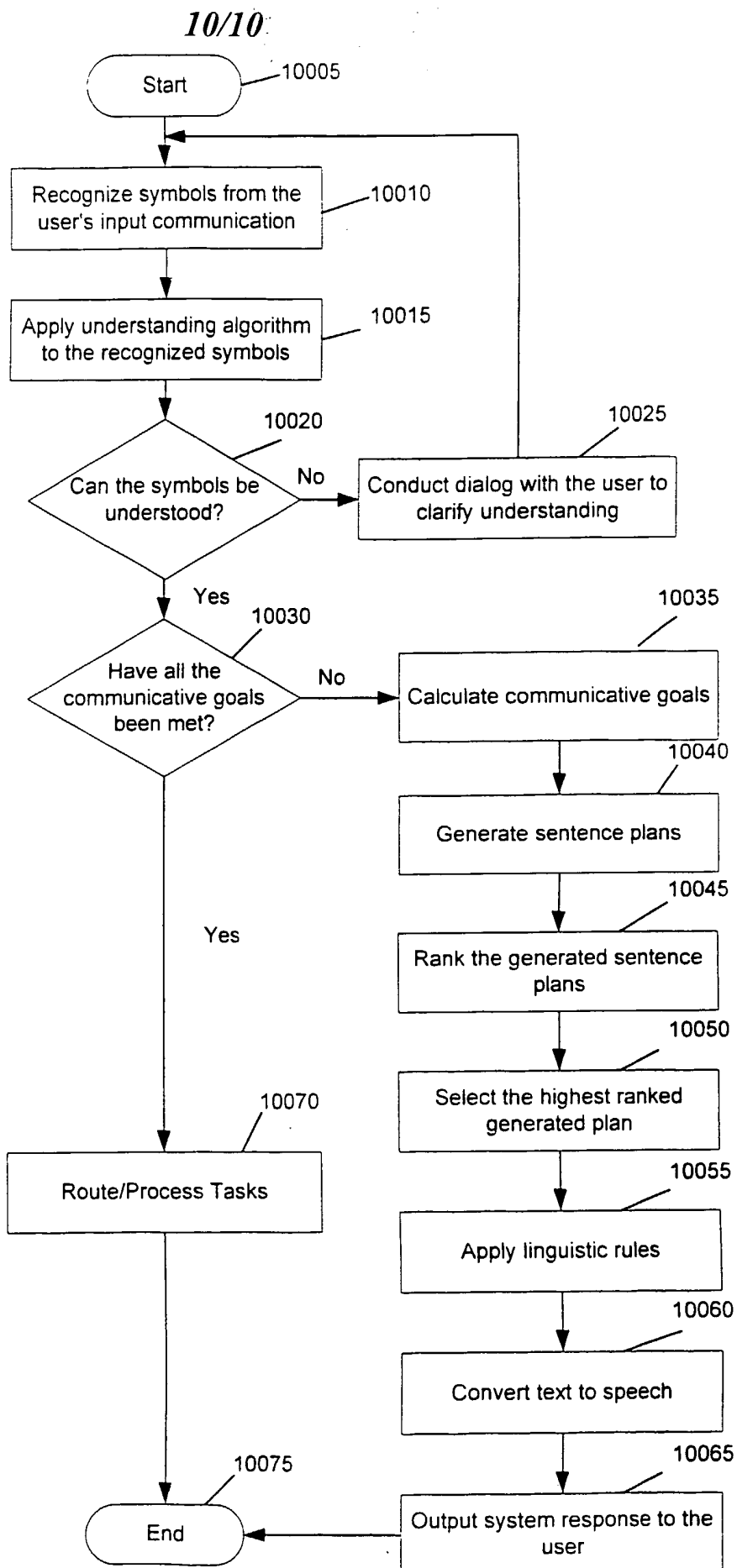


Fig. 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07235

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 17/20, 17, 28; G10L 5/06, 9/00

US CL : 704/1, 9, 10, 270, 275

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/1, 9, 10, 270, 275

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
west/east**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,675,707 A (GORIN ET AL) 07 OCTOBER 1997, Abstract	1-27
Y	US 5,357,596 A (TAKEBAYASHI ET AL) 18 OCTOBER 1994, Abstract, figure 40, col. 28, line 1 through col. 31, line 2.	1-27
A	US 5,860,063 A (GORIN ET AL) 12 JANUARY 1999, Abstract.	1-27
Y	US 6,035,267 A (WATANABE ET AL) 07 MARCH 2000, Abstract, figures 5-14, col. 4, line 65 through col. 11, line 28.	1-27



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"G" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

17 MAY 2002

Date of mailing of the international search report

01 JUL 2002

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

PATRICK N. EDOUARD

Telephone No. (703) 308-6725