US 20070122792A1

(19) United States

(12) Patent Application Publication

Galley et al.

(10) Pub. No.: US 2007/0122792 A1

(43) Pub. Date: May 31, 2007

(54) LANGUAGE CAPABILITY ASSESSMENT AND TRAINING APPARATUS AND TECHNIQUES

(76) Inventors: **Michel Galley**, New York, NY (US); **Kevin Knight**, Marina Del Rey, CA (US); **Daniel Marcu**, Hermosa Beach, CA (US)

Correspondence Address:
CARR & FERRELL LLP
2200 GENG ROAD
PALO ALTO, CA 94303 (US)

(21) Appl. No.: **11/272,460**

(22) Filed: **Nov. 9, 2005**

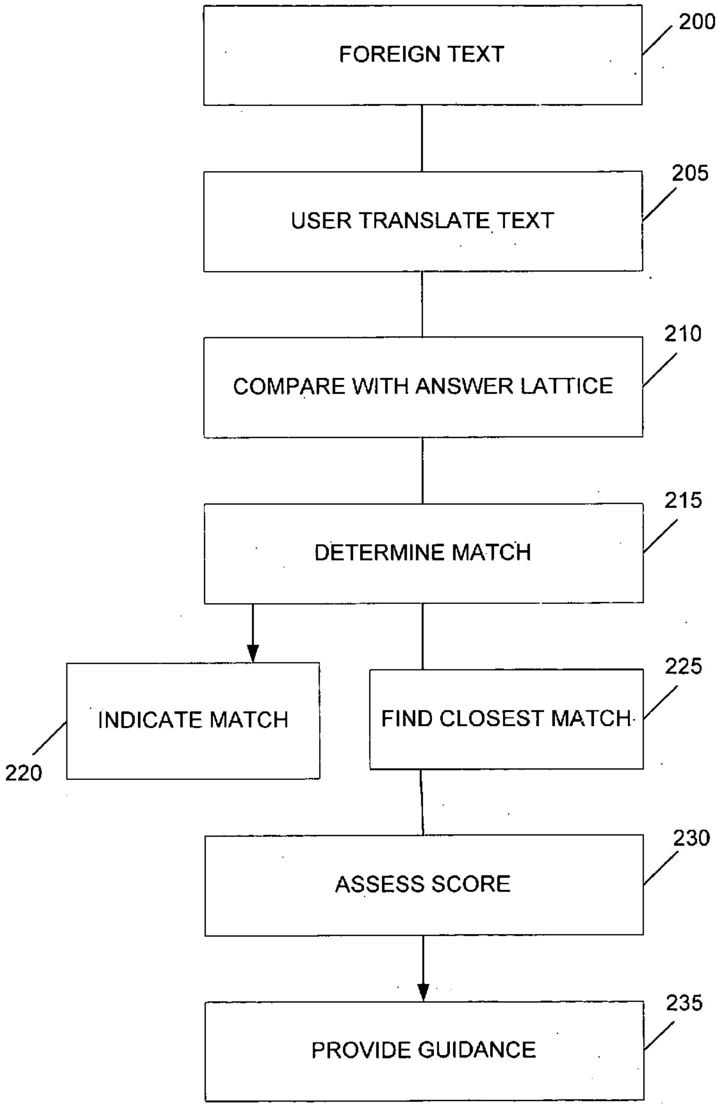
Publication Classification

(51) **Int. Cl.**
G09B 7/00 (2006.01)

(52) **U.S. Cl.** **434/353**

(57) **ABSTRACT**

A learning system for a text-to-text application such as a machine translation system. The system has questions, and a matrix of correct answers to those questions. Any of the many different correct answers within the matrix can be considered as perfectly correct answers to the question. The system operates by displaying a question, which may be a phrase to be translated, and obtaining an answer to the question from the user. The answer is compared against the matrix and scored. Feedback may also be provided to the user.



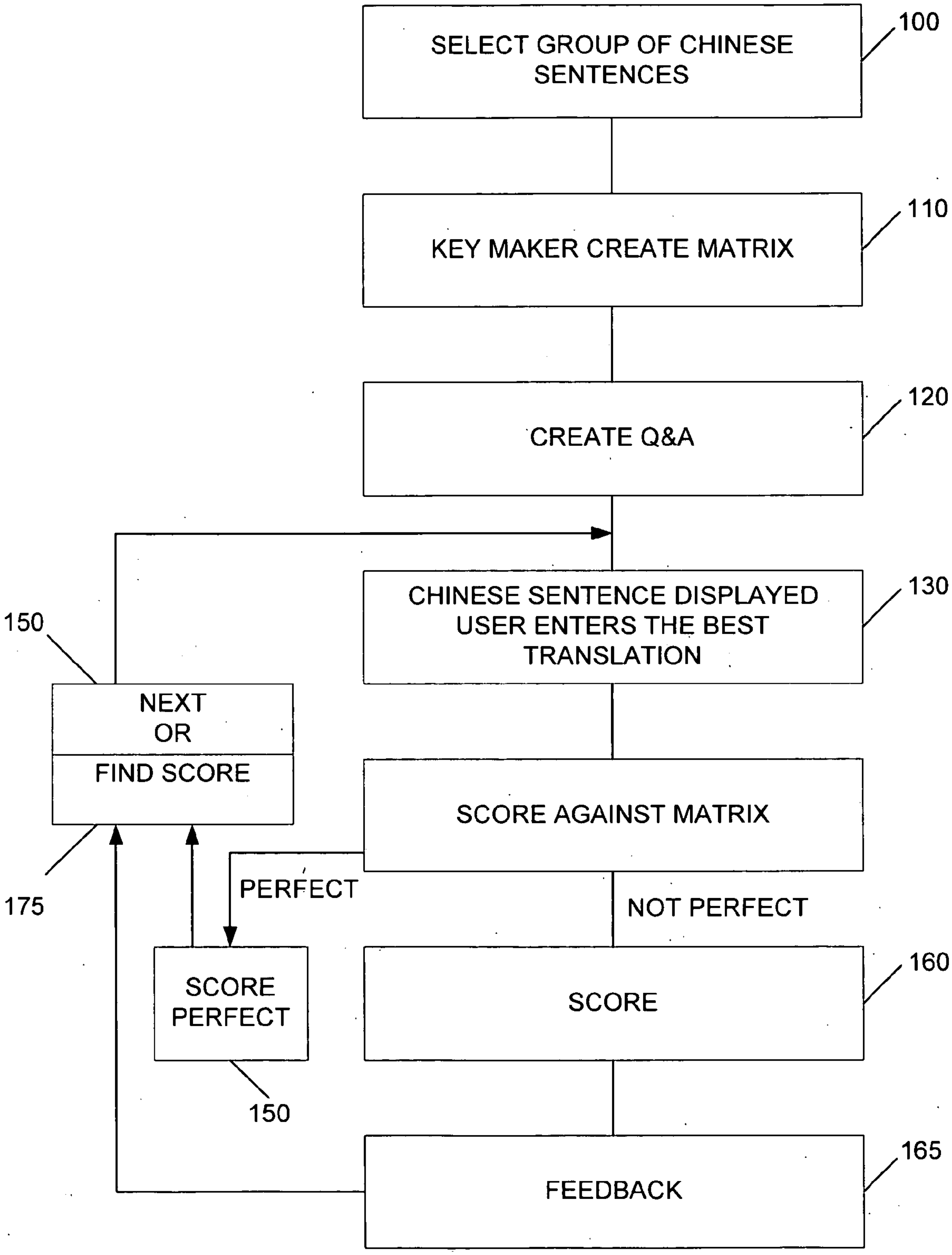

SELECT GROUP OF CHINESE SENTENCES
100
KEY MAKER CREATE MATRIX
110
CREATE Q&A
120
CHINESE SENTENCE DISPLAYED USER ENTERS THE BEST TRANSLATION
130
150
NEXT OR
FIND SCORE
175
SCORE AGAINST MATRIX
PERFECT
NOT PERFECT
SCORE PERFECT
150
SCORE
160
FEEDBACK
165

FIG. 1

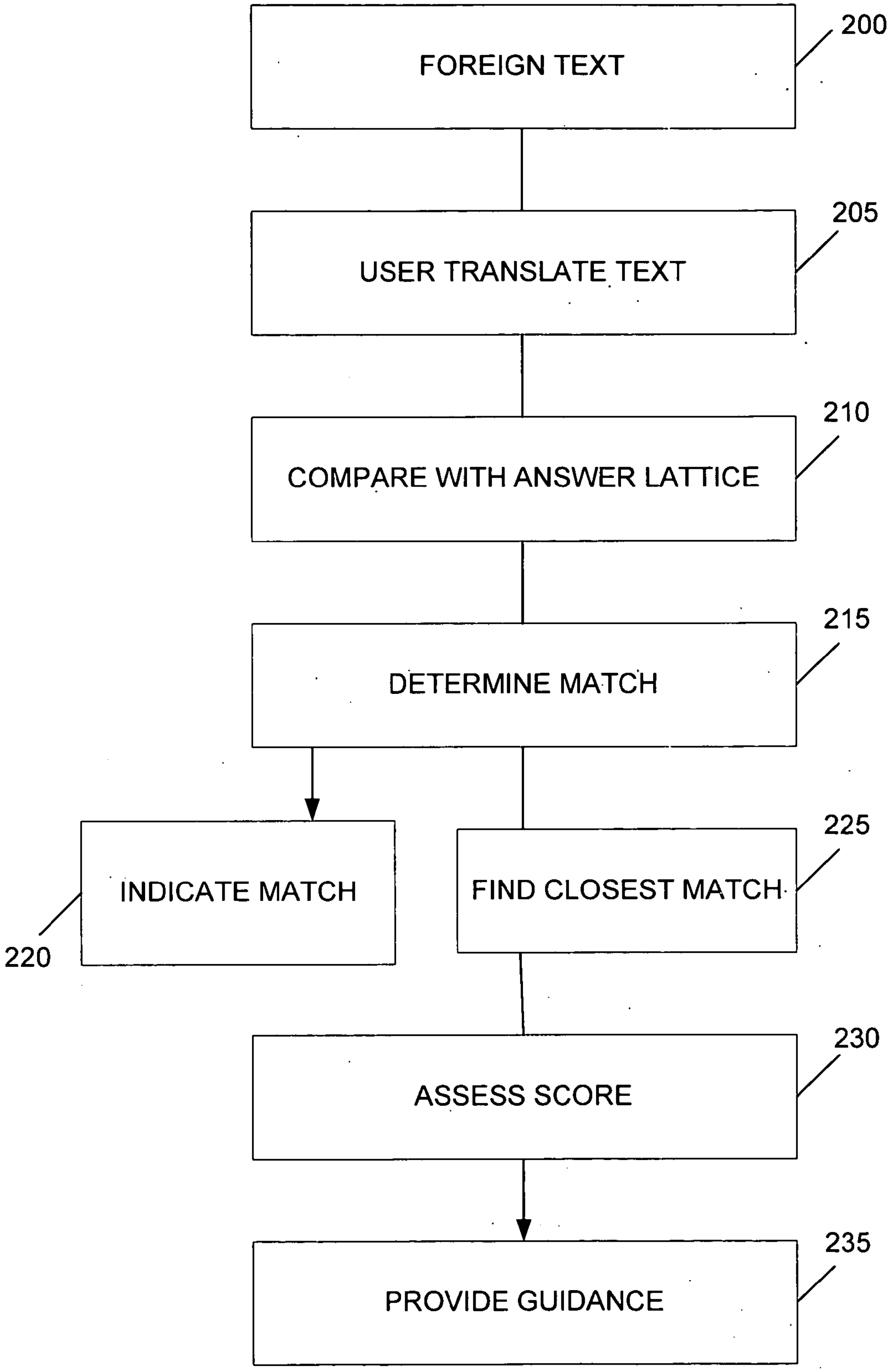

FOREIGN TEXT
200
USER TRANSLATE TEXT
205
COMPARE WITH ANSWER LATTICE
210
DETERMINE MATCH
215
INDICATE MATCH
220
FIND CLOSEST MATCH
225
ASSESS SCORE
230
PROVIDE GUIDANCE
235

FIG. 2

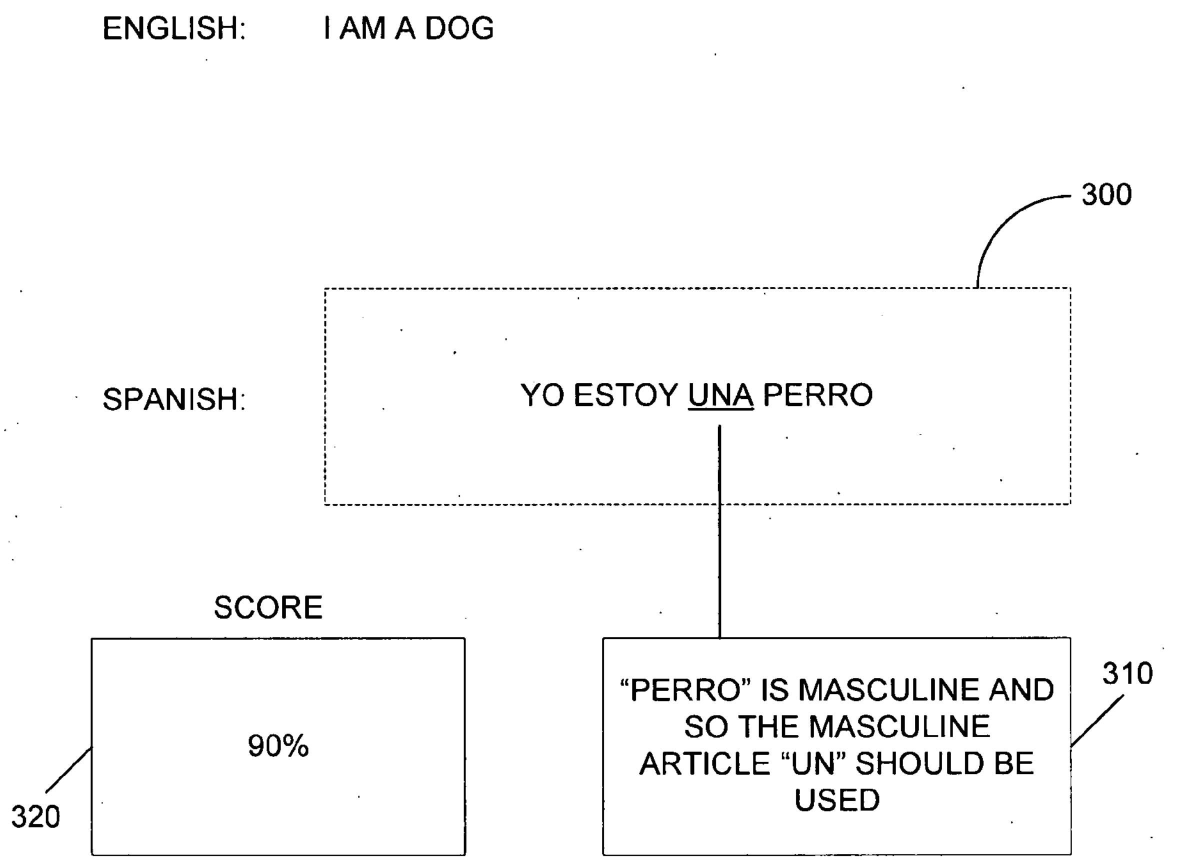
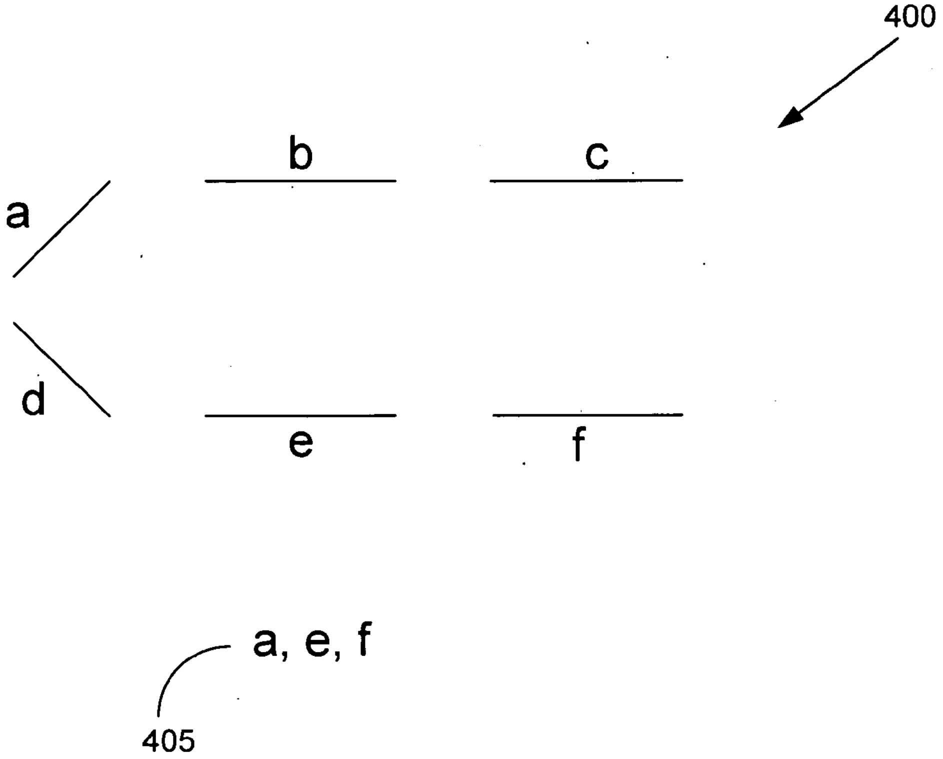

ENGLISH: I AM A DOG
300
SPANISH:
YO ESTOY UNA PERRO
SCORE
90%
320
"PERRO" IS MASCULINE AND SO THE MASCULINE ARTICLE "UN" SHOULD BE USED
310

FIG. 3



	Input	Path	Cost
411-	a	a	0
412-	a	d	1
413-	ae	ab	1
414-	a	ab	1
415-	ae	de	1
416-	aef	abc	2
417-	aef	def	1

FIG. 4

LANGUAGE CAPABILITY ASSESSMENT AND TRAINING APPARATUS AND TECHNIQUES

BACKGROUND

[0001] Text-to-text applications may be used for various purposes, including speech recognition, machine translation from one language to another, as well as automated summarization. A typical text to text application learns information from a training corpus, and uses the learned information to carry out the text to text operation.

[0002] One text to text application is machine translation, which is often used to automatically translate from one language to another. Machines including computers have also been used for educational purposes, such as in classrooms and the like.

SUMMARY

[0003] The present application describes a new text to text application which allows assessing a users ability to translate from a first language into a second language. According to an aspect, the application is used for matching an entered answer against a correct answer, and producing an output based on training data within the text to text application, where there are many different correct answers, each of which is completely correct.

[0004] An aspect includes that there may be many correct answers, since, for example, there may be many ways of translating phrases from the source language to the target language. In an embodiment, any answer that is entered by the user is compared against an entire matrix of correct answers.

[0005] Another aspect describes providing feedback to the user indicating their mistakes, and providing at least one helping them identify those mistakes.

[0006] An embodiment describes the text to text application as being language translation, and in the embodiment, the application helps the user to learn a new language by assessing their abilities. An embodiment provides feedback which can be used as part of the learning tool. The feedback may provide more detailed information about which parts of their abilities are lacking and/or better answers.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] These and other aspects will now be described with reference to the accompanying drawings, wherein:

[0008] FIG. **1** shows a flowchart of the operation of the system;

[0009] FIG. **2** shows a flowchart of operation of the overall operation of the system;

[0010] FIG. **3** illustrates a user screen; and

[0011] FIG. **4** illustrates a speed up technique.

DETAILED DESCRIPTION

[0012] The general structure and techniques, and more specific embodiments which can be used to effect different ways of carrying out the more general goals are described herein.

[0013] The overall operation of the embodiment is illustrated with reference to FIG. **1**. FIG. **1** illustrates how this system would operate for testing a person's translation ability. It should be understood, however, that this system is also capable of being used for determining speech ability, or determining summarization ability, or more generally, for analyzing any text-to-text application where there are a number of different possible correct answers. The system operates by comparing an input from a user to the plurality of possible correct answers.

[0014] For the embodiment, the system is assessing a person's Chinese to English translation ability. For example, a government agency may want to hire people who can translate Chinese documents into fluent English. At the same time, the system may provide tools to help individual translators improve their Chinese to English skills. The two basic tasks of assessment and tutoring may be provided to different users, or to the same user. It should of course be understood that any language pair can be used in place of the Chinese and English, and that Chinese and English are provided as being an exemplary language pair.

[0015] In operation, first a number of test questions must be selected for assessment. At **100**, a group of Chinese sentences is selected. This may be done manually and placed into a database, or may be done automatically by simply choosing sentences from a database.

[0016] At **110**, a key maker is used to build a network of correct English translations. The key maker may be a person that manually builds the translation.

[0017] Alternatively, the key maker may use a machine that automatically translates, with final checking for example done by a person. A special user interface is used so that the key maker can facilitate the processing. The final network has millions of correct translation, for example. There are likely millions of possible correct answers for any translations.

[0018] **110** shows this generically as creating a matrix, but it should be understood that different ways of representing the multiple different answers that are contemplated in this embodiment.

[0019] At **120**, the operation of the program begins, by creating questions and answers "Q&A" and creating a user interface that is accessible by a human.

[0020] The operations described herein may be carried out on computers, which may be any kind of computer, either general purpose, or some specific purpose computer such as a workstation. The computer may be a Pentium class computer, running Windows XP or Linux, or may be a McIntosh computer. The programs may be written in C, or Java, or any other programming language. The programs may be resident on a storage medium, e.g., magnetic or optical, e.g. the computer hard drive, a removable disk or other removable medium. The programs may also be run over a network, for example, with a server or other machine sending signals to the local machine, which allows the local machine to carry out the operations described herein.

[0021] FIG. **2** shows an overall flow diagram of this system when used for detecting correct answers in a language translation context. The system is intended to be used to help learn a new language, and to assess end-user ability in that language.

[0022] In operation, **200** represents the foreign language text being displayed to a user or student. The user translates the text at **205**. The translation is compared with a matrix of prestored correct answers at **210**. There may be millions of correct ways of translating any foreign language phrase or sentence into another language. In the embodiment, these many different ways of translating are represented in a compact form, where certain common paths are not restored. The translation is compared, and at **215** a match is determined. If an exact match is determined, **220** indicates that by indicating that an exact match has been determined. If no exact match is determined, then the closest match is determined at **225**, and a score is assessed at **230**. The score may represent the extent to which the correct answer deviated from the given answer. At **235**, the program may produce a feedback guidance screen that indicates information about what errors the user made in the translation, and how to fix them. The feedback is based on the specific kind of error that was made by the user.

[0023] In an embodiment, the correct answer is stored in a compact graphical representation where paths through the graph may represent many different possible correct answers. The compact representation may be a recursive transition network ("RTN"), in which a graph is represented with certain parts of the graph represented by variables that represent commonly occurring portions within the graph. Another alternative may include representing the correct answer using IDL or weighted IDL.

[0024] The comparison between the user's entered answer and the lattice of answers determines one of a number of different kinds of errors which can exist. Exemplary errors may include word insertion (an extra word being inserted), word deletion (a missing word), word substitution (the wrong word being used), word permutation (wrong order to words), word stemming (wrong or different endings to the words), and paraphrasing (similar meaning but not exactly the same meaning). In order to obtain an accurate assessment of the degree of error, each of these may be appropriately analyzed. In one aspect, each generic error is counted as a single error, even if it causes many different word and/or placement variations. For example, an embodiment counts a word permutation as a single error even though two different words are out of order. In addition, the different errors, such as word insertion and word deletion may be counted as different values; for example, insertion may be less seriously weighted than deletion.

[0025] **235** represents providing feedback based on the specific kind of error. The feedback may be displayed or otherwise provided to the user. In one embodiment, the feedback may be displayed as a display which shows the errors and some possible ways to deal with the errors. For example, this may show a display such as shown in FIG. **3**. FIG. **3** shows the sample sentence "I am a dog" in English, being translated by the student in the text box **300**. The sample output with the incorrect words (here "una") being underlined, and a brief description **310** of the kind of error which has been noted. A score is also provided as **320**, which represents the number of errors noted in the translation, and the kinds of errors.

[0026] In one aspect, an answer key of all the answers is provided. This can be provided in the form for example of a graph. Nodes in the graph which represent synonyms that may be present at various places in the graph may be replaced by shorthand representations of those synonyms, such as a variable. For example, if the sentence is about a battle, then battle, fight and fighting may be synonyms which may be present in the graph. Each place where those words are duplicated will be replaced by a single transition, for example a transition labeled as "A". A special graphical user interface may be used for making this graph. The graphical user interface is basically a drawing program that generates the different nodes in the graph, but also allows those nodes to be translated into the answer lattice when complete. Another aspect is that the GUI may generate random sentences as paths through the graph, to enable testing the graph. Another aspect is that the GUI may enable minimization of the graph, that is to remove duplicate parts of the graph, by replacing those duplicate parts by the variables that represent commonly used transitions.

[0027] In operation, the model answer is compared against the different aspects in the graph, to determine "costs". Different kinds of errors may lead to different kinds of costs, depending on how important those errors are. The costs can be set by trial and error, or can be simply assigned. There may be a lower cost for insertion or deletion of pronouns, and compared that to a higher cost for use of the noun, e.g., the use of the word Apple versus Orange.

[0028] The overall algorithm may simply use a brute force approach which exhaustively searches through the graph. However, this may not be practical in terms of processing power, since it may require analyzing each of the perhaps millions of correct answers for a text-to-text system.

[0029] A speed up technique is described with reference to FIG. **4** the algorithm speed of technique which uses a search over a complete estimator. For example, at each step, a few hypotheses may be maintained. Each hypothesis includes a current cost, and the characteristic. All of the different paths are always maintained, but the path with the most promising hypothesis is followed. A few of the different search states may also be maintained. Only one of the search states will have the lowest cost, and a heuristic between the remaining length and the end of the finite state search length may also be used. By assessing the promise of each state, and being conservative with the estimate, it is ensured that the estimate will always be in the correct position.

[0030] The speed up technique will be illustrated in the following. FIG. **4** shows the State graph **400**, and the input to the State graph. The input is shown as **405**. A priority queue **410** is maintained which represents the different states as the system passes through the queue.

[0031] For the first pass through the queue, two entries are created shown as **411** and **412**. The first entry is shown as the input, here a, the path here also a, and the cost, here 0. The second input **412** includes the input, here a, the path, here d, and the cost which here is 1. The path **411** has the minimum cost, so successor states to that path are created as **413**, **414**. However, the path **412** remains. The successor states shown as **413**, **414**, here ae/ab with a cost of 1, and a/ab, with a cost of 1. Eventually, the state aef/def is reached with cost **1**, which is the best match for aef.

[0032] Each of these pieces represent a representation of a position in the lattice. This produces a stack of states that can be searched backwards.

[0033] The above has described this being used for translation, however it should be understood that the same techniques can also be applied to summarization, speech recognition, or testing of pronunciation. Any kind of language problem that has multiple answers can be handled in this way.

[0034] The feedback can simply be feedback which is tailored to specific errors, or alternatively can be feedback which is quoted back from a grammar book, or from the translation database. Queries into the database looking for specific languages that are targeted to the example can be used.

[0035] In another aspect, a specific error that is made of can be represented by a special type of indicia, and specific and special guidance for that, and error can be provided. According to another aspect, the feedback can represent all of the legal ways to say for specific thing.

[0036] Although only a few embodiments have been disclosed in detail above, other embodiments are possible and the inventor (s) intend these to be encompassed within this specification. The specification describes specific examples to accomplish a more general goal that may be accomplished in other way. This disclosure is intended to be exemplary, and the claims are intended to cover any modification or alternative which might be predictable to a person having ordinary skill in the art. For example, other applications of this system may be possible.

[0037] Also, the inventor(s) intend that only those claims which use the words "means for" are intended to be interpreted under 35 USC 112, sixth paragraph. Moreover, no limitations from the specification are intended to be read into any claims, unless those limitations are expressly included in the claims.

What is claimed is:

1. A method comprising:

accepting input from a user representing the user's answer to a text to text based question, where there are many different correct answers, each of which is completely correct; and

accessing a database which includes said many different correct answers in a packed format where at least a plurality of said answers use common information to represent parts of said answers; and

determining an error between said input and any of said correct answers.

2. A method as in claim 1, wherein said determining an error comprises scoring the answers according to a kind of said error.

3. A method as in claim 1, wherein said scoring the answers comprises counting an error only once even when multiple parts are affected by said error.

4. A method as in claim 1, wherein said text to text based question comprises a question about translation between a first language and a second language.

5. A method as in claim 1, wherein said text to text based question comprises pronunciation of speech.

6. A method as in claim 1, further comprising determining, from said error, feedback to help the user in understanding more about said error.

7. A method as in claim 1, further comprising producing an answer matrix which includes each of said many different correct answers, prior to said accessing.

8. A method as in claim 7, wherein said producing an answer matrix comprises forming a user interface which facilitates entry of the many different answers.

9. A method as in claim 8, wherein said user interface includes an associated tool that tests certain ones of said many different answers.

10. A method as in claim 1, wherein said packed format is a recursive translation network.

11. A method as in claim 1, further comprising analyzing said answer by maintaining a list including most prominent single hypotheses through the packed database, and following the most promising hypothesis.

12. A method as in claim 3, wherein said errors include word insertion errors, Word deletion errors, word substitution errors, word permutation errors, word stemming errors, and paraphrasing errors.

13. A method comprising:

displaying text in a first language to a user;

accepting input of the text in a second language from the users as a translation;

comparing said translation to a matrix of prestored answers;

determining either an exact match for a closest match between the translation and any answer in the matrix; and

providing a score indicative of either the closest match or the exact match.

14. A method as in claim 13, further comprising providing feedback to the user indicating information about errors made in said translation.

15. A method as in claim 14, wherein said providing a score comprises scoring each of a plurality of errors as being a single error even when said error affects more than one word.

16. A method as in claim 15, wherein said scoring comprises continued considering a single error for each of the word insertion, word deletion, word substitution, word permutation, word stemming, and paraphrasing.

17. A method as in claim 11, wherein said determining comprises following a most promising path through the matrix of correct answers.

18. An apparatus comprising:

an interface part operating to accept input from a user representing the user's answer to a text to text based question, where there are many different correct answers, each of which is completely correct;

a database which includes said many different correct answers in a packed format, where at least a plurality of said answers rely on common information for parts of said answers; and

a machine, which operates to determine an error between said input and said many different correct answers by identifying a closest correct answer to said input, and identifying an error between said closest answer and said input.

19. An apparatus as in claim 18, wherein said machine scores the answers according to a kind of said error.

20. An apparatus method as in claim 18, wherein said machine scores the answers comprises counting an error only once even when multiple parts are affected by said error.

21. An apparatus as in claim 18, wherein said text to text based question comprises a question about translation between a first language and a second language.

22. An apparatus as in claim 18, wherein said text to text based question comprises pronunciation of speech.

23. An apparatus as in claim 18, wherein said machine determines, from said error, feedback to help the user in understanding more about said error.

24. An apparatus as in claim 18, wherein said machine produces an answer matrix which includes each of said many different correct answers, prior to said accessing.

25. An apparatus as in claim 24, wherein said machine operates to form a answer matrix user interface that allows entry of details of said answer matrix, and which facilitates entry of the many different answers.

26. An apparatus as in claim 25, wherein said answer matrix user interface includes an associated tool that tests certain ones of said many different answers.

27. An apparatus as in claim 18, wherein said packed format is a recursive translation network.

28. An apparatus as in claim 18, wherein said machine operates by analyzing said answer by maintaining a list including a most prominent single hypotheses through the packed database, and following the most promising hypothesis.

29. An apparatus as in claim 20, wherein said errors include word insertion errors, Word deletion errors, word substitution errors, word permutation errors, word stemming errors, and paraphrasing errors.

30. An apparatus comprising:

a machine that includes a memory storing a matrix of prestored answers to questions, said machine operating to produce signals indicative of a user interface that operates to display text in a first language to a user, and accepts input of the text in a second language from the user as a translation, said machine determining either an exact match for a closest match between the translation and any answer in the matrix, and providing a score indicative of a relationship between said input of said text, and any of said answers.

31. An apparatus as in claim 30, wherein said machine further provides feedback to the user indicating information about errors made in said translation.

32. An apparatus as in claim 14, wherein said score scores each of a plurality of errors as being a single error even when said error affects more than one word.

* * * * *