US 20050055209A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2005/0055209 A1**

Epstein et al. (43) **Pub. Date: Mar. 10, 2005**

(54) **SEMANTIC LANGUAGE MODELING AND CONFIDENCE MEASUREMENT**

(76) Inventors: **Mark E. Epstein**, Katonah, NY (US); **Hakan Erdogan**, Istanbul (TR); **Yuqing Gao**, Mount Kisco, NY (US); **Michael A. Picheny**, White Plains, NY (US); **Ruhi Sarikaya**, Shrub Oak, NY (US)

Correspondence Address:
KEUSEY, TUTUNJIAN & BITETTO, P.C.
14 VANDERVENTER AVENUE, SUITE 128
PORT WASHINGTON, NY 11050 (US)

(21) Appl. No.: **10/655,838**

(22) Filed: **Sep. 5, 2003**

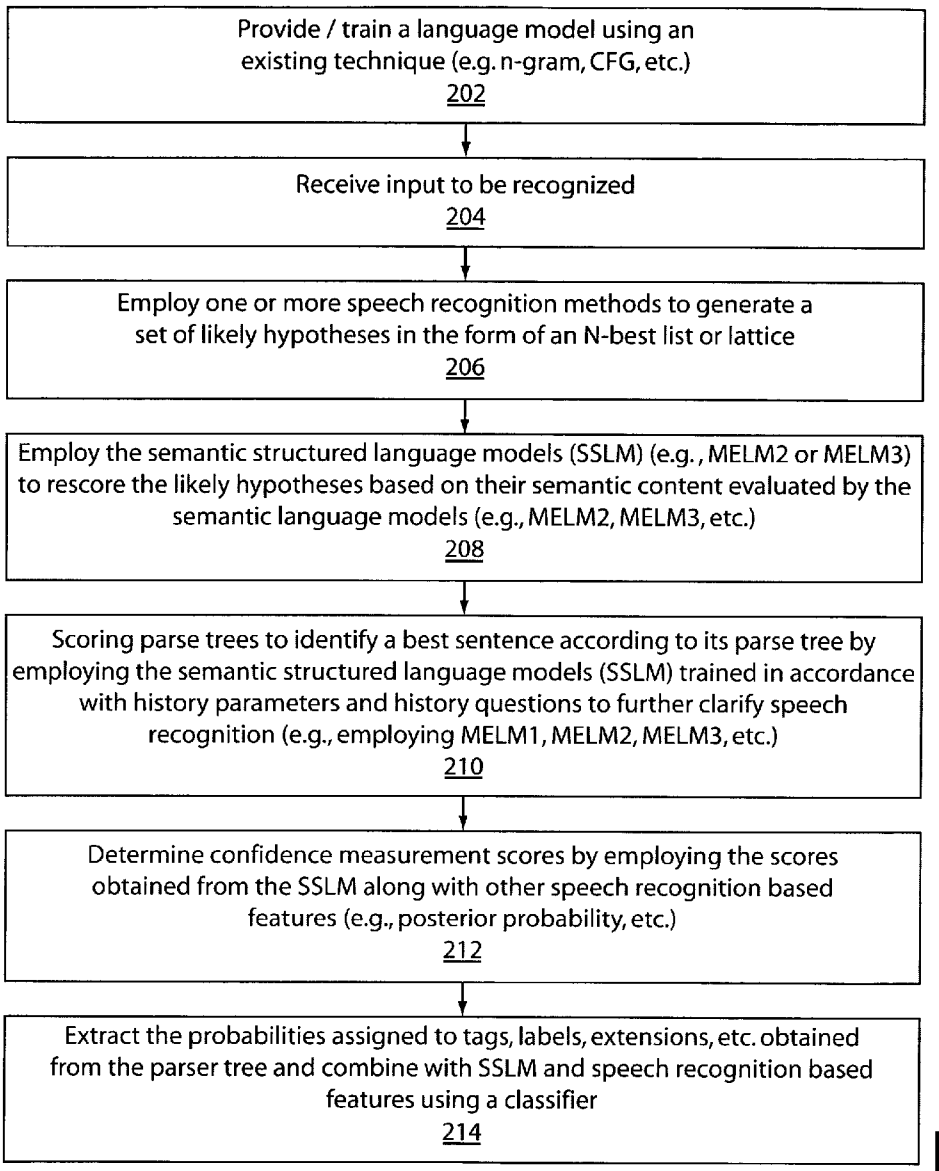
Publication Classification

(51) **Int. Cl.**[7] **G10L 15/00**

(52) **U.S. Cl.** **704/255**

(57) **ABSTRACT**

A system and method for speech recognition includes generating a set of likely hypotheses in recognizing speech, rescoring the likely hypotheses by using semantic content by employing semantic structured language models, and scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech.



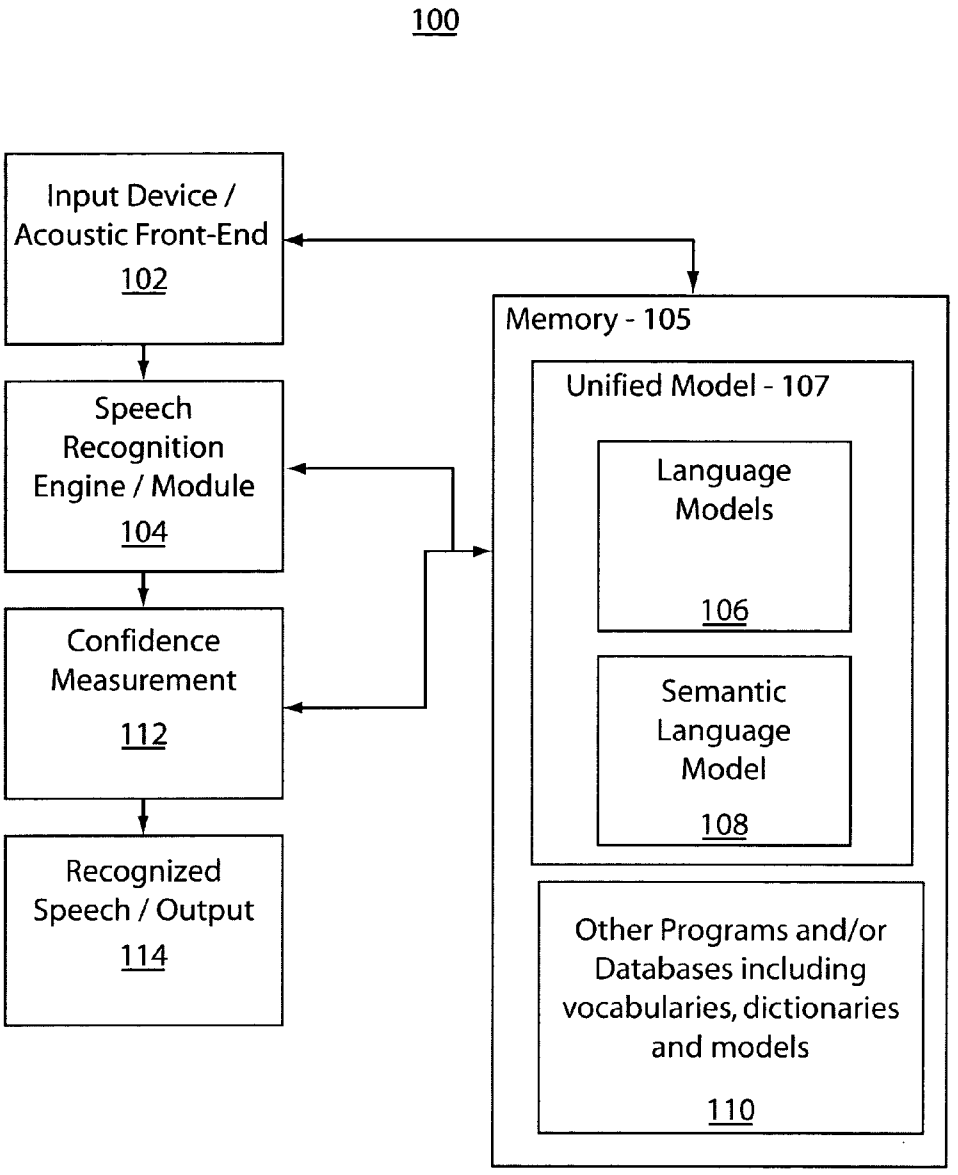
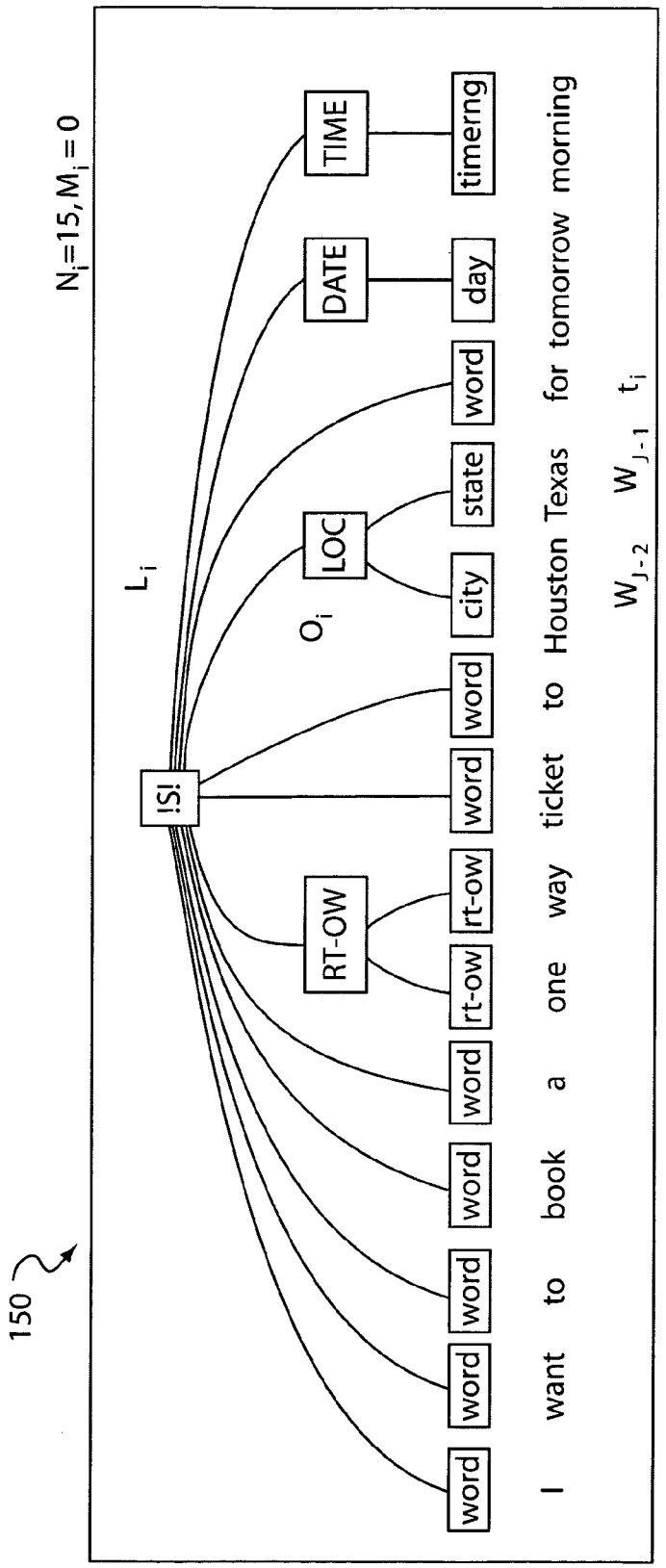

100
Input Device / Acoustic Front-End 102
Speech Recognition Engine / Module 104
Confidence Measurement 112
Recognized Speech / Output 114
Memory - 105
Unified Model - 107
Language Models 106
Semantic Language Model 108
Other Programs and/or Databases including vocabularies, dictionaries and models 110

FIG. 1



[!S! i want to book a [RT-OW one way RT-OW] ticket to [LOC Houston Texas LOC] for [DATE tomorrow DATE] [TIME morning TIME]!S!]

FIG. 2

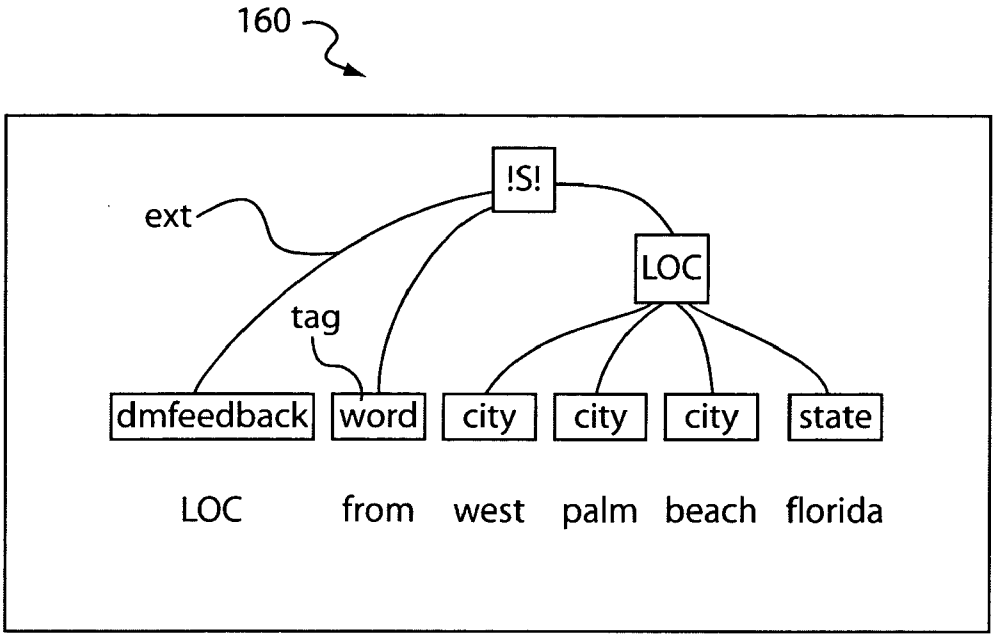

160
!S!
ext
LOC
tag
dmfeedback
word
city
city
city
state
LOC
from
west
palm
beach
florida

FIG. 3

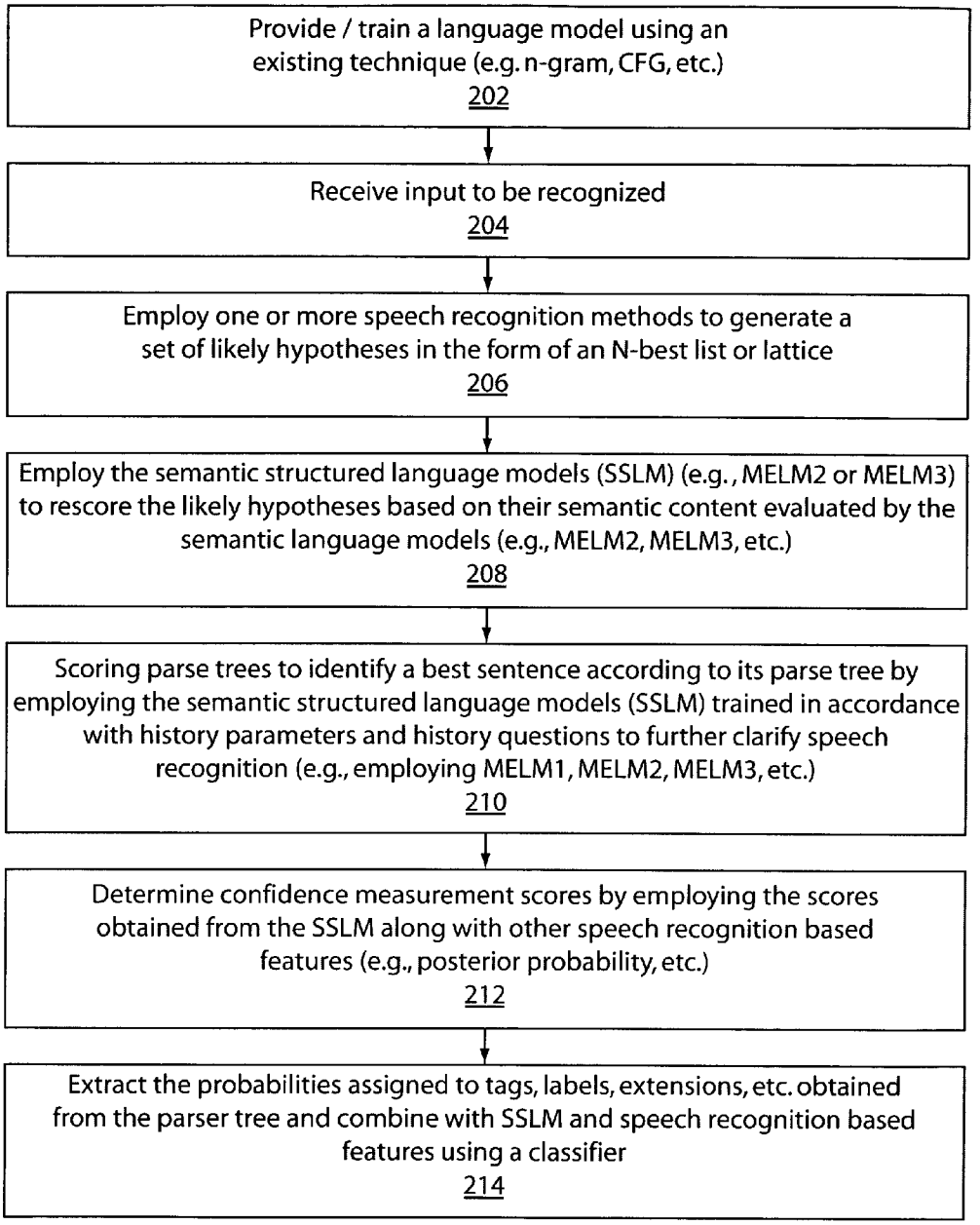

Provide / train a language model using an existing technique (e.g. n-gram, CFG, etc.)
202
Receive input to be recognized
204
Employ one or more speech recognition methods to generate a set of likely hypotheses in the form of an N-best list or lattice
206
Employ the semantic structured language models (SSLM) (e.g., MELM2 or MELM3) to rescore the likely hypotheses based on their semantic content evaluated by the semantic language models (e.g., MELM2, MELM3, etc.)
208
Scoring parse trees to identify a best sentence according to its parse tree by employing the semantic structured language models (SSLM) trained in accordance with history parameters and history questions to further clarify speech recognition (e.g., employing MELM1, MELM2, MELM3, etc.)
210
Determine confidence measurement scores by employing the scores obtained from the SSLM along with other speech recognition based features (e.g., posterior probability, etc.)
212
Extract the probabilities assigned to tags, labels, extensions, etc. obtained from the parser tree and combine with SSLM and speech recognition based features using a classifier
214

FIG. 4

SEMANTIC LANGUAGE MODELING AND CONFIDENCE MEASUREMENT

BACKGROUND

[0001] 1. Field of Exemplary Embodiments

[0002] Aspects of the invention relate to language modeling, and more particularly to systems and methods which use semantic parse trees for language modeling and confidence measurement.

[0003] 2. Description of the Related Art

[0004] Large vocabulary continuous speech recognition (LVCSR) often employs statistical language modeling techniques to improve recognition performance. Language modeling provides an estimate for the probability of a word sequence (or sentence) P(w1 w2 w3 . . . wN) in a language or a subdomain of a language. A prominent method in statistical language modeling is n-gram language modeling, which is based on estimating the sentence probability by combining probabilities of each word in the context of previous n−1 words.

[0005] Although n-gram language models achieve a certain level of performance, they are not optimal. N-grams do not model the long-range dependencies, semantic and syntactic structure of a sentence accurately.

[0006] A related problem to modeling semantic information in a sentence is the confidence measurement based on semantic analysis. As the speech recognition output will always be subject to some level of uncertainty, it may be vital to employ some measure that indicates the reliability of the correctness of the hypothesized words. The majority of approaches to confidence annotation methods use two basic steps: (1) generate as many features as possible based on speech recognition and/or a natural language understanding process, (2) use a classifier to combine these features in a reasonable way.

[0007] There are a number of overlapping speech recognition based features that are exploited in many studies (see e.g., R. San-Segundo, B. Pellom, K. Hacioglu and W. Ward, "Confidence Measures for Spoken Dialog Systems", ICASSP-2001, pp. 393-396, Salt Lake City, Utah, May 2001; R. Zhang and A. Rudnicky, "Word Level Confidence Annotation Using Combination of Features", Eurospeech-2001, Aalborg, Denmark, September, 2002; and C. Pao, P. Schmid and J. Glass, "Confidence Scoring for Speech Understanding Systems", ICSLP-98, Sydney, Australia, December 1998). For domain independent large vocabulary speech recognition systems, posterior probability based on a word graph is shown to be the single most useful confidence feature (see, F. Wessel, K. Macherey and H. Ney, "A Comparison of Word Graph and N-best List Based Confidence Measures", pp.1587-1590, ICASSP-2000, Istanbul, Turkey, June 2000). Semantic information can be considered as an additional information source complementing speech recognition information. In many, if not all, of the previous studies the way the semantic information is incorporated into the decision process is rather ad hoc. For example in C. Pao et al., "Confidence Scoring for Speech Understanding Systems", referenced above, the semantic weights assigned to words are based on heuristics. Similarly, in P. Carpenter, C. Jin, D. Wilson, R. Zhang, D. Bohus and A. Rudnicky, "Is This Conversation on Track", Eurospeech-2001, pp. 2121-2124, Aalborg, Denmark, September 2001, such semantic features as "uncovered word percentage", "gap number", "slot number", etc. are generated experimentally in an effort to incorporate semantic information into the confidence metric.

SUMMARY

[0008] A system and method for speech recognition, includes a unified language model including a semantic language model and a lexical language model. A recognition engine finds a parse tree to analyze a word group using the lexical model and the semantic models. The parse tree is selected based on lexical information and semantic information, which considers tags, labels, and extensions to recognize speech.

[0009] Preferred methods may be integrated into a speech recognition engine or applied to lattices or N-best lists generated by speech recognition.

[0010] A method for speech recognition includes generating a set of likely hypotheses in recognizing speech, rescoring the likely hypotheses by using semantic content by employing semantic structured language models, and scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech.

[0011] In other embodiments, the step of determining a confidence measurement is included. The confidence measurement determination may include includes employing a statistical method to combine word sequences with a parser tree to determine a confidence score for recognized speech. This may include determining the confidence measurement by employing scores obtained from the semantic structured language models along with other speech recognition based features. The scores may be obtained by extracting probabilities assigned to tags, labels and extensions obtained from a parser tree. The step of combining the semantic structured language models and speech recognition based features with the extracted probabilities using a classifier may be included

[0012] These and other objects, features and advantages of the present exemplary systems and methods will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

[0013] The exemplary embodiments will be described in detail in the following description of preferred embodiments with reference to the following figures wherein:

[0014] **FIG. 1** is a block diagram showing a speech recognition and confidence measurement system in accordance with the present disclosure;

[0015] **FIG. 2** is a diagram showing an illustrative parse tree employed to recognize speech and further shows information (e.g., w_{j-2}, w_{j-1}, L_i, O_i, N_i and M_i) obtained from the parse tree to build a semantic language model in accordance with the present disclosure;

[0016] **FIG. 3** is a diagram showing an illustrative classer tree with probabilities assigned employed to provide confidence scores in accordance with the present disclosure; and

[0017] **FIG. 4** is a block/flow diagram showing a speech recognition and confidence measurement method in accordance with the present disclosure.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0018] The present disclosure provides a system and method, which incorporates semantic information in a semantic parse tree into language modeling. The semantic structured language modeling (SSLM) methods may employ varying levels of lexical and semantic information using any of the statistical learning techniques including, for example, maximum entropy modeling, decision trees, neural networks, support vector machines or simple counts. In one embodiment, maximum entropy modeling is used. This embodiment will be employed as an illustrative example herein.

[0019] In accordance with this disclosure, a set of methods is based on semantic analysis of sentences. These techniques utilize information extracted from parsed sentences to statistically model semantic and lexical content of the sentences. A maximum entropy method is employed, for example, to rescore N-best speech recognizer hypotheses using semantic features in addition to lexical features.

[0020] The maximum entropy method (MEM) may be used for language modeling in the context of n-grams, sentence-based statistical language modeling and syntactic structured language models. However, an exemplary embodiment of the present disclosure employs MEM to incorporate semantic features into a unified language model. This integration enables one to easily use semantic features in language modeling. Semantic features can be obtained from a statistical parser as well as from a stochastic recursive transition network (SRTN). These features encode information related to the semantic interpretation of each word and word groups, which is one important consideration to distinguish meaningful word sequences from less meaningful or meaningless ones.

[0021] It should be understood that the elements shown in the FIGS. may be implemented in various forms of hardware, software or combinations thereof. Preferably, these elements are implemented in software on one or more appropriately programmed general-purpose digital computers having a processor and memory and input/output interfaces.

[0022] Referring now to the drawings in which like numerals represent the same or similar elements and initially to **FIG. 1**, a system **100** for carrying out one embodiment is shown. System **100** may include a computer device or network, which provides speech recognition capabilities. System **100** may be employed to train speech recognition models or may be employed as a speech recognition system or speech analyzer. System **100** may include an input device for inputting text or speech to be recognized. Input device **102** may include a microphone, a keyboard, a mouse, a touch screen display, a disk drive or any other input device. Inputs to input device **102** may include semantic information, which can be stored in memory **105**. This semantic information may be employed to construct semantic language models **108** in accordance with the present disclosure.

[0023] Semantic information is employed in a semantic parse tree to be used in language modeling. The semantic structured language modeling (SSLM) methods or programs stored in language models **106** and semantic language model **108** may employ varying levels of lexical and semantic information using any of the statistical learning techniques including, for example, maximum entropy modeling, decision trees, neural networks, support vector machines or simple counts. This method may be carried out be employing a speech recognition engine or engines **104**.

[0024] Speech recognition module **104** processes acoustic input, digital signals, text or other input information to recognize speech or organize the input to create language models **106**, semantic language models **108** or other programs or databases **110**, such as vocabularies, dictionaries, other language models or data, etc. Engine **104** may provide speech recognition capabilities for system **100** by employing any known speech recognition methods.

[0025] In accordance with this disclosure, engine **104** also employs a set of methods based on semantic analysis of sentences to enhance the speech recognition capabilities of system **100**. These techniques utilize information extracted from parsed sentences to statistically model semantic and lexical content of the sentences. A maximum entropy method is employed, for example, to rescore N-best speech recognizer hypotheses using semantic features in addition to lexical features.

[0026] The maximum entropy method (MEM) may be used for language modeling in the context of n-grams, sentence-based statistical language modeling and syntactic structured language models. MEM is illustratively employed as an exemplary embodiment in the present disclosure to incorporate semantic features into a unified language model **107**. This integration enables easy usage of semantic features in language modeling. Semantic features can be obtained from a statistical parser as well as from a stochastic recursive transition network (SRTN), which may be incorporated in module **104**.

[0027] These parsing features encode information related to the semantic interpretation of each word and word groups, which is one important consideration to distinguish meaningful word sequences from less meaningful or meaningless ones.

[0028] Semantic information may include one or more of word choice, order of words, proximity to other related words, idiomatic expressions or any other information based word, tag, label, extension or token history.

[0029] This semantic information is employed to develop language modeling by training a model **108** based on the semantic information available. These models may include the following illustrative semantic structured language models (SSLM).

[0030] MELM1 (Maximum Entropy Method 1) preferably uses unigram, bigram and trigram features. It is possible to use more intelligent features that will capture a longer range and higher-level information. Considering data sparsity and computation requirements, the following sublist of context question types for individual token probability computations may be employed (see MELM2).

[0031] MELM2 (Maximum Entropy Method 2) uses longer-range semantic features, for example, 7 types of features:

[0032] Unigram: (default question)

[0033] bigram: previous word wj−1 (ignore label tokens)

[0034] trigram: two previous words wj−1 and wj−2 (ignore label tokens)

[0035] Current active parent label Li (parent constituent label)

[0036] Ni (number of tokens to the left since current L starts)

[0037] Oi (previous closed constituent label)

[0038] Mi (number of tokens to the left after Oi finishes)

[0039] 7 types of questions: (default), (wj−1), (wj−1, wj−2), (Li), (Li, Ni), (Li, Ni, wj−1), (Oi, Mi)

[0040] Note that, these are the questions are chosen for the maximum entropy (ME) model. There may be many other possible features that utilize other information such as tags, grandparent labels etc. The choices could be dependent on the domain or the type of semantic parsing employed. The maximum entropy framework enables one to incorporate any type of features as long as they are computable.

[0041] The model **108** is employed to calculate word probabilities to decipher probabilities that particular word sequences or phrases have been employed.

[0042] Referring to **FIG. 2**, an example application of MELM2 to compute P(ti=for| history) is presented as a parser tree.

[0043] The probability of the token sequence, [!S! I want to book a [RT-OW one way RT-OW] ticket to [LOC Houston Tex. LOC] for [DATE tomorrow DATE][TIME morning TIME] is equivalent to joint probability of a classer tree and the word sequence given as the following equation:

$$P(W,C)\approx\pi_{ni-1}P(t_i|t_1,\ldots,t_i-3,t_i-2,t_i-1)$$

[0044] Where a token t can be a word, label, tag, etc.

[0045] Another SSLM includes MELM3 (Maximum Entropy Method 3), which combines semantic classer and parser and uses a full parse tree **150**. The full parse tree **150** presents a complete semantic structure of the sentence where, in addition to classer information, such as RT-OW (round-trip one way), LOC (location), DATE, TIME, semantic relationships between the constituents are also derived, e.g., wj−2, wj−1, t_i. The following features are used to train a Maximum Entropy based statistical model:

[0046] 7 history parameters of MELM3

[0047] wj−1: previous word wj−1 (ignore label tokens)

[0048] wj−2: previous word of previous word (ignore label tokens)

[0049] L: (parent constituent label)

[0050] N: (number of tokens to the left since L starts)

[0051] O: (previous closed constituent label)

[0052] M: (number of tokens to the left after O finishes)

[0053] G: (grandparent label)

[0054] 6 history question types: (default), (wj−1), (wj−1, wj−2), (L,N), (O,M), (L,G)

[0055] Although the trees that the questions are based on are different, MELM2 and MELM3 share similar questions. Indeed, only the fifth question of MELM3 is not included in the MELM2 question set. Note that even though these specific question sets are selected for MELM2 and MELM3, any question based on classer and parser trees can be a legitimate choice in Maximum Entropy modeling.

[0056] The inventors experimentally determined that these question sets performed adequately in training a language model. Inclusion of additional questions did not significantly improve the performance for the illustratively described task.

[0057] The set of semantic language modeling techniques, MELM1, MELM2 and MELM3 improve speech recognition accuracy. In addition, features derived from these language models can be used for confidence measurement by employing confidence measurement module **112** (**FIG. 1**).

[0058] Module **112** uses the language model score for a given word in MELM2 model, which is conditioned not only on previous words but also tags, labels and relative coverage of these labels over words. Tags define the types of words regarding their semantic content. For example, the tag assigned to Houston is "city". Words that are not semantically important are assigned a "null" tag. Labels are used to categorize a word or word group into one of the concepts. The number of labels is less than the number of tags. An extension or arc is the connection between a tag assigned to a word and the label. Relative coverage refers to how far the current word is from the beginning of the current label.

[0059] MELM2 presents an effective statistical method to combine word sequences with a semantic parse tree. Therefore, the MELM2 score, for example, may be used as a feature for confidence measurement. However, MELM2 for a given word only depends on the previous word sequence and the parse tree up to that word. A low score can be expected for the current word if the previous word is recognized incorrectly. Besides the MELM2 score for the current word wi, a window of three words ([w{i−1} wi w{i+1}]), MELM2-ctx3, were considered and five words, MELM2-ctx5, centered on the current word to capture the context information. The same features can be derived for MELM3 as well.

[0060] Referring to **FIG. 3** with continued reference to **FIG. 1**), probabilities obtained by module **104** from semantic parse trees stored in model **108** can also be used for confidence measurement in module **112**. The classer/parser performs a left-to-right bottom-up search to find the best parse tree for a given sentence. During the search, each tag node (tag), label node (e.g., LOC) and extension (ext) in the parse tree is assigned a probability. Similarly, an extension probability represents the probability of placing that extension between the current node and its parent node given the "context". When the parser is conducting the search both lexical (from model **106**) and the semantic clues (from model **108**) are used to generate the best parser action.

[0061] The degree of confidence while assigning the tag and the label feature values is reflected in the associated probabilities. If the word does not "fit" in the current lexical and semantic context, its tag and labels are likely to be

assigned low probabilities. Therefore, using these probabilities as features in the confidence measurement is a viable way to capture the semantics of a sentence. Below is the classer tree for the phrase "from West Palm Beach Fla.". The corresponding classer tree is shown in **FIG. 3**. cTag (shown as "tag" in **FIG. 3**) and cTagExt (shown as "ext" in **FIG. 3**) are classer tag and tag extension probabilities, respectively. Likewise in a parser tree, as opposed to a classer tree, "arc" and "ext" would correspond to pTag and pTagExt, which are parser tag and tag extension probabilities, respectively.

[**0062**] A classer tree **160** is shown in **FIG. 3** along with its text representation. Each token has a pair of probabilities.

[**0063**] {0.502155 {!S!_1_1:LOC_dmfeedback_1_ 0.997937 from_word_0.99371_0.995734 {LOC_ 0.999976_0.998174 west_city_0.635543_0.894638 palm_city_0.998609_0.981378 beach_city_0.998609_ 0.957721 florida_state_0.96017_0.995701 LOC_ 0.999976_0.998174}!S!_1_1}}

[**0064**] Confidence measurements by module **112** are optional, but can greatly improve speech recognition accuracy, which is output in block **114**. Output **114** may be customized to any form. For example, output **114** may be speech synthesized and acoustically rendered, textually render, transmitted as an analog or digital signal, etc.

[**0065**] The following are some specific examples where SSLM corrects errors committed by regular n-gram methods. These examples were run by the inventors to show the advantages of the present embodiment using MEM. Confidence scores are also given below to show improvements.

Reference (Ref):	new horizons and and blue chip
n-gram:	log is an end blue chip
sslm:	horizons and and blue chip
Ref:	what is my balance by money type
n-gram:	what was the of my money sent
sslm:	what is my balance by money take
Ref:	change pin
n-gram:	change plan
sslm:	change pin

[**0066**] The following are some specific examples where errors committed by posterior probability features are corrected by the semantic confidence features. The threshold for confidence is set to 0.87, which roughly corresponds to 5% False Acceptance rate for both posterior probability (post) and sslm+post (sslm and posterior probability). These examples are correctly accepted by sslm+post features but falsely rejected by the post features alone:

Ref:	balance
Hypothesis (Hyp):	balance
Post:	0.54 (confidence measure)
Post + sslm:	0.95 (confidence measure)
Ref:	summary
Hyp:	summary
Post:	0.63 (confidence measure)
Post + sslm:	0.93 (confidence measure)
Ref:	plan
Hyp:	plan

-continued

Post:	0.79 (confidence measure)
Post + sslm:	0.88 (confidence measure)

[**0067**] The following examples are correctly rejected by sslm+post features but falsely accepted with post features alone:

Ref:	call
Hyp:	<SIL>
post_conf:	0.88 (confidence measure)
post + sslm:	0.04 (confidence measure)
Ref:	—
Hyp:	have
post_conf:	0.93 (confidence measure)
post + sslm:	0.82 (confidence measure)
Ref:	representative
Hyp:	rep
post_conf:	0.88 (confidence measure)
post + sslm:	0.70 (confidence measure)

[**0068**] Referring to **FIG. 4**, a method for speech recognition includes providing or training a language model in block **202**. The language model may be trained using known training techniques, such as n-gram, CFG, etc. In block **204**, input to be recognized is received. The input may be in any form permitted by the speech recognition method or device. In block **206**, one or more speech recognition methods may be employed to generate a set of likely hypotheses. The hypotheses are preferably in the form of an N-best list or lattice structure.

[**0069**] In block **208**, semantic structured language models (SSLM) are employed to rescore the likely hypotheses based on the semantic content of the hypotheses. This is performed by evaluating the hypothesis using the SSLM models, e.g., MELM2 or MELM3, etc. In block **210**, parse trees are scored to identify a best sentence in accordance with its parse tree. This is performed by using SSLMs trained in accordance with history parameters and history questions to further clarify the speech recognized.

[**0070**] The history parameters may include a previous word (wj−1), a previous word of the previous word (wj−2), a parent constituent label (L), a number of tokens (N) to the left since L starts, a previous closed constituent label (O), a number of tokens (M) to the left after O finishes, and a grandparent label (G). The history questions may include a default, (wj−1), (wj−1, wj−2), (L,N), (O,M), and (L,G).

[**0071**] In block **212**, a confidence measurement or score may be determined by employing the scores obtained from the SSLM along with other speech recognition based features, e.g., posterior probability, etc. In block **214** probabilities assigned to tags, labels, extensions, etc. obtained from the parser tree may be combined with SSLM and speech recognition based features using a classifer. These probabilities may be employed to further improve speech recognition by increasing the level of confidence in confidence scores.

[**0072**] Having described preferred embodiments for semantic language modeling and confidence measurement (which are intended to be illustrative and not limiting), it is noted that modifications and variations can be made by

persons skilled in the art in light of the above teachings. It is therefore to be understood that changes may be made in the particular embodiments disclosed which are within the scope and spirit of the present disclosure as outlined by the appended claims. Having thus described the exemplary embodiments with the details and particularity required by the patent laws, what is claimed and desired protected by Letters Patent is set forth in the appended claims.

What is claimed is:

1. A method for speech recognition, comprising the steps of:

generating a set of likely hypotheses in recognizing speech;

rescoring the likely hypotheses by using semantic content by employing semantic structured language models; and

scoring parse trees to identify a best sentence according to the sentences' parse tree by employing the semantic structured language models to clarify the recognized speech.

2. The method as recited in claim 1, further comprising the step of training a language model using speech recognition methods.

3. The method as recited in claim 1, wherein the set of likely hypotheses is in the form of an N-best list or lattice.

4. The method as recited in claim 1, wherein the step of rescoring employs MELM2 or MELM3 semantic structured language models.

5. The method as recited in claim 1, wherein the step of scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech includes the step of training the structured semantic language models in accordance with history parameters and history questions.

6. The method as recited in claim 5, wherein the history parameters include a previous word (wj−1), a previous word of the previous word (wj−2), a parent constituent label (L), a number of tokens (N) to the left since L starts, a previous closed constituent label (O), a number of tokens (M) to the left after O finishes, and a grandparent label (G).

7. The method as recited in claim 5, wherein the history questions include a default, (wj−1), (wj−1, wj−2), (L,N), (O,M), and (L,G).

8. The method as recited in claim 1, further comprising the step of determining a confidence measurement.

9. The method as recited in claim 8, wherein the step of determining a confidence measurement includes employing a statistical method to combine word sequences with a parser tree to determine a confidence score for recognized speech.

10. The method as recited in claim 8, wherein the step of determining a confidence measurement includes employing scores obtained from the semantic structured language models along with other speech recognition based features.

11. The method as recited in claim 1, further comprising the step of extracting probabilities assigned to tags, labels and extensions obtained from a parser tree.

12. The method as recited in claim 11, further comprising the step of combining the semantic structured -language models and speech recognition based features with the extracted probabilities using a classifier.

13. The method as recited in claim 1, wherein the semantic structured language models are trained by employing unigram, bigram and trigram features.

14. The method as recited in claim 1, wherein the semantic structured language models are trained using one or more of relative labels, token numbers, and answers to questions related to word order or placement.

15. The method as recited in claim 1, wherein the semantic structured language models are trained by including a unigram feature, a bigram feature, a trigram feature, a current active parent label (Li), a number of tokens (Ni) to the left since current parent label (Li) starts, a previous closed constituent label (Oi), a number of tokens (Mi) to the left after the previous closed constituent label finishes, and a number of questions to classify parser tree entries.

16. The method as recited in claim 15, wherein the questions include a default, (wj−1), (wj−1, wj−2), (Li), (Li, Ni), (Li, Ni, wj−1), and (Oi, Mi), where w represents a word and j is and index representing word position.

17. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for speech recognition, in accordance with claim 1.

18. A method for speech recognition, comprising the steps of:

generating a set of likely hypotheses in recognizing speech;

rescoring the likely hypotheses by using semantic content by employing semantic structured language models; and

scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech; and

determining a confidence measurement by employing scores obtained from the semantic structured language models along with other speech recognition based features.

19. The method as recited in claim 18, wherein the set of likely hypotheses is in the form of an N-best list or lattice.

20. The method as recited in claim 18, wherein the step of rescoring employs MELM2 or MELM3 semantic structured language models.

21. The method as recited in claim 18, wherein the step of scoring parse trees to identify a best sentence according to the sentence's parse tree by employing the semantic structured language models to clarify the recognized speech includes the step of training the semantic structured language models in accordance with history parameters and history questions.

22. The method as recited in claim 21, wherein the history parameters include a previous word (wj−1), a previous word of the previous word (wj−2), a parent constituent label (L), a number of tokens (N) to the left since L starts, a previous closed constituent label (O), a number of tokens (M) to the left after O finishes, and a grandparent label (G).

23. The method as recited in claim 21, wherein the history questions include a default, (wj−1), (wj−1, wj−2), (L,N), (O,M), and (L,G).

24. The method as recited in claim 18, further comprising the step of extracting probabilities assigned to tags, labels and extensions obtained from a parser tree.

25. The method as recited in claim 24, further comprising the step of combining the semantic structured language models and speech recognition based features with the extracted probabilities using a classifier.

26. The method as recited in claim 18, wherein the semantic structured language models are trained by employing unigram, bigram and trigram features.

27. The method as recited in claim 18, wherein the semantic structured language models are trained using one or more of relative labels, token numbers, and answers to questions related to word order or placement.

28. The method as recited in claim 18, wherein the semantic structured language models are trained by including a unigram feature, a bigram feature, a trigram feature, a current active parent label (Li), a number of tokens (Ni) to the left since current parent label (Li) starts, a previous closed constituent label (Oi), a number of tokens (Mi) to the left after the previous closed constituent label finishes, and a number of questions to classify parser tree entries.

29. The method as recited in claim 28, wherein the questions include a default, (wj−1), (wj−1, wj−2), (Li), (Li, Ni), (Li, Ni, wj−1), and (Oi, Mi), where w represents a word and j is and index representing word position.

30. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for speech recognition, in accordance with claim 18.

31. A system for speech recognition, comprising:

a unified language model including a semantic language model and a lexical language model;

a recognition engine which finds a parse tree to analyze a word group using the lexical model and the semantic models, wherein the parse tree is selected based on lexical information and semantic information which considers tags, labels, and extensions to recognize speech.

32. The system as recited in claim 31, wherein the parser tree includes semantic information and classer information used in identifying a best parser tree for a given word group.

33. The system as recited in claim 31, wherein the parser tree includes information extracted from parsed sentences to statistically model semantic and lexical content of sentences.

34. The system as recited in claim 31, wherein the semantic language model includes unigram, bigram and trigram features.

35. The system as recited in claim 31, wherein the semantic language model includes one or more of relative labels, token numbers, and answers to questions related to word order or placement.

36. The system as recited in claim 31, wherein the semantic model is trained by including a unigram feature, a bigram feature, a trigram feature, a current active parent label (Li), a number of tokens (Ni) to the left since current parent label (Li) starts, a previous closed constituent label (Oi), a number of tokens (Mi) to the left after the previous closed constituent label finishes, and a number of questions to classify parse tree entries.

37. The system as recited in claim 36, wherein the questions include a default, (wj−1), (wj−1, wj−2), (Li), (Li, Ni), (Li, Ni, wj−1), and (Oi, Mi), where w represents a word and j is and index representing word position.

38. The system as recited in claim 31, wherein the semantic model is trained by including history parameters and history questions.

39. The system as recited in claim 38 wherein the history parameters include a previous word (wj−1), a previous word of the previous word (wj−2), a parent constituent label (L), a number of tokens (N) to the left since L starts, a previous closed constituent label (O), a number of tokens (M) to the left after O finishes, and a grandparent label (G).

40. The system as recited in claim 39, wherein the history questions include a default, (wj−1), (wj−1, wj−2), (L,N), (O,M), and (L,G).

41. The system as recited in claim 31, further comprising a confidence measurement module.

42. The system as recited in claim 31, wherein the confidence measurement module employs a statistical method to combine word sequences with the parse tree to determine a confidence score for recognized speech.

43. The system as recited in claim 31, wherein the confidence measurement module extracts probabilities assigned to tag nodes, label nodes and extensions in the parse tree.

* * * * *