

圖 1

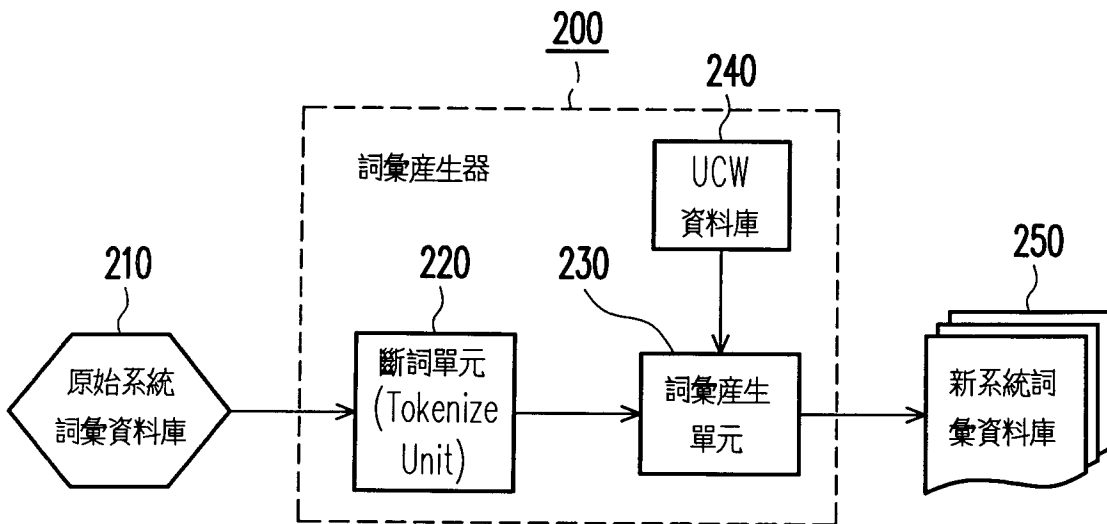


圖 2



發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號： 94102410

※申請日期： 94. 1. 27

※IPC 分類： G06F 17/27 (2006.01)

G06L 15/02 (2006.01)

一、發明名稱：(中文/英文)

詞彙產生方法、裝置及運用詞彙產生裝置之語音
辨識系統/Vocabulary Generating Apparatus and
Method thereof and Speech Recognition System
with the Vocabulary Generating Apparatus

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

台達電子工業股份有限公司/DELTA ELECTRONICS, INC.

代表人：(中文/英文) 鄭崇華/BRUCE C. H. CHENG

住居所或營業所地址：(中文/英文)

桃園縣龜山工業區興邦路 31-1 號/NO. 31-1, SHIEN BAN RD., KUEI
SAN INDUSTRIAL ZONE, TAOYUAN HSIEN, TAIWAN, R. O. C.

國 籍：(中文/英文) 中華民國/TW

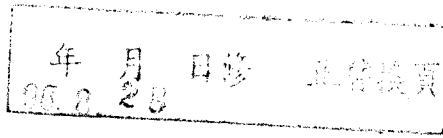
三、發明人：(共 2 人)

姓 名：(中文/英文)

1 蔡錦和/ Ching-Ho Tsai

2 黃良聲/ Liang-Sheng Huang

國 籍：(中文/英文) 中華民國/TW



四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。



九、發明說明：

【發明所屬之技術領域】

本發明是關於一種語音辨識系統及其詞彙產生的方法與裝置。特別是關於一種詞彙產生方法、裝置及運用上述詞彙產生裝置之語音辨識系統，此詞彙產生方法可增加語音辨識系統的彈性，使得使用者在不能準確的說出系統命令時，仍可以使用“未知的代稱”來代替命令中不能確定的部分。

【先前技術】

語音辨識系統可隨著技術之精進，而讓使用者非常方便地使用相關之設備，例如操作電視、音響等等電子設備器材。使用者只要根據系統所設定可辨識之語音指令，就可以非常方便地操作這些設備。除了運用在電氣設備外，語音辨識之技術亦已廣泛地運用在相關的領域中，例如身份辨識、語音輸入等等之運用。

然而，現行的語音辨識系統最常見的一個問題，在於使用者面對語音辨識系統時，往往不知道或是忘記如何對語音辨識系統下命令。然而，語音辨識系統所能辨識的範圍，僅限於系統設計時所預定下的詞彙以及語法。因此，使用者如無法確定系統所能接受的詞彙，便無法享受語音系統的便利。

【發明內容】

本發明提出一個詞彙產生方法、裝置及運用上述詞彙產生裝置之語音辨識系統。此詞彙產生方法可增

加語音辨識系統的彈性，使得使用者在不能準確的說出系統命令時，可以使用本發明特別定義的一種“未知的代稱”(Unknown Code Word)來代替命令中不能確定的部分。

為達上述之目的，本發明提出一種詞彙產生裝置，包括斷詞單元、未知的代稱資料庫與詞彙產生單元。此斷詞單元用以接收多數個原始系統詞彙，並進行斷詞(Tokenize)操作，並根據每個原始系統詞彙產生對應之多數含有意義之單位。此未知的代稱資料庫用以儲存多數個未知的代稱。此詞彙產生單元連接到斷詞單元與未知的代稱資料庫，用以接收上述含有意義之單位以及未知的代稱，並進行詞彙合併操作，以便將斷詞單元所產生上述之含有意義之單位結合未知的代稱而合成多數個新系統詞彙。

上述之詞彙產生裝置，其中斷詞單元根據每個原始系統詞彙產生對應之含有意義之單位為多數個字或詞。

上述之詞彙產生裝置，其中原始系統詞彙是來自一語音辨識系統所原有的系統詞彙資料庫內之多數個系統詞彙。

上述之詞彙產生裝置，其中詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是捨棄部分之含有意義之單位而以部分之未知的代稱所代替。

上述之詞彙產生裝置，其中詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位之部分捨棄，並加入未知的代稱所代替。

上述之詞彙產生裝置，其中詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位直接加入一個或多個未知的代稱。

上述之詞彙產生裝置，其中詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位先行重組後直接加入未知的代稱而產生新系統詞彙。上述之詞彙產生裝置，在一實施例中，亦可省略部分有意義之單位後再進行重組，甚至加入未知的代稱以產生新的系統詞彙。

為達上述之目的，本發明提出一種詞彙產生方法，適用於一語音辨識系統，而此語音辨識系統具有一原始系統詞彙資料庫。此方法包括將原始系統詞彙資料庫內的多數個原始系統詞彙進行斷詞(Tokenize)操作，並根據每個原始系統詞彙產生對應之多數含有意義之單位。而後將含有意義之單位以及未知的代稱進行詞彙合併操作，以便將斷詞單元所產生含有意義之單位結合未知的代稱而合成多數個新系統詞彙，以提供語音辨識系統用以進行語音辨識。

上述的詞彙產生方法，其中根據每個原始系統詞彙產生對應之含有意義之單位為多數個字或詞。

在一實施例中，上述的詞彙產生方法，其中根據含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是捨棄部分之含有意義之單位而以未知的代稱所代替。或是將含有意義之單位之部分捨棄，並加入未知的代稱所代替。在一實施例中，是將含有意義之單位直接加入該未知的代稱。

在一實施例中，是將含有意義之單位直接加入多數個該些未知的代稱。在另一實施例中，是將含有意義之單位先行重組後直接加入一個或多個未知的代稱而產生新系統詞彙。

為達上述之目的，本發明提出一種具有詞彙產生裝置之語音辨識系統，包括原始系統詞彙資料庫、未知的代稱資料庫、斷詞單元與詞彙產生單元。此原始系統詞彙資料庫用以儲存多數個原始系統詞彙。此未知的代稱資料庫用以儲存多數個未知的代稱。此斷詞單元，用以接收上述之原始系統詞彙，並進行斷詞(Tokenize)操作，並根據每個原始系統詞彙產生對應之多數含有意義之單位。而詞彙產生單元連接到斷詞單元與未知的代稱資料庫，用以接收含有意義之單位以及未知的代稱，並進行詞彙合併操作，以便將斷詞單元所產生含有意義之單位結合未知的代稱而合成多數個新系統詞彙。

上述之語音辨識系統，在一實施例中，其詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是捨棄部分之含有意義之單位而以未知的代稱所代替。

上述之語音辨識系統，在一實施例中，其詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位之部分捨棄，並加入未知的代稱所代替。

上述之語音辨識系統，在一實施例中，其詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位直接加入一個或多個未知的代稱。

上述之語音辨識系統，在一實施例中，其詞彙產生單元將接收含有意義之單位以及未知的代稱進行詞彙合併操作，以產生新系統詞彙是將含有意義之單位先行重組後直接加入未知的代稱而產生新系統詞彙。

【實施方式】

本發明提出一個詞彙產生方法、裝置及運用上述詞彙產生裝置之語音辨識系統。此詞彙產生方法可增加語音辨識系統的彈性，使得使用者在不能準確的說出系統命令時，可以使用本發明特別定義的一種“未知的代稱”(Unknown Code Word)來代替命令中不能確定的部分。

語音辨識系統的主要流程如圖 1 所示，在語者輸入一段語音訊號後，語音辨識系統 100 經由一語音特徵值擷取單元 110 擷取出此語音訊號的特徵值 (Features)。而後，然後根據此語音辨識系統 100 的詞彙特徵模型比對單元 120 對此語音訊號所擷取出來的特徵值進行詞彙模型的比對，而此比對是根據一詞彙資料庫 130 取得系統詞彙及其特徵模型後，由詞彙特徵模型比對單元 120 根據詞彙取得相近的模型，並輸出至辨識決定單元 140，依照既定的規則 (Rule) 得到對應於此語音訊號的辨識詞句 (Recognized Text) 之辨識結果。

所謂系統的『詞彙特徵模型』，乃是由系統預定辨識的每一個『詞彙〈Vocabulary〉』的『特徵模型』所組成。因此，當語者輸入的語音訊號片段的特徵值和某一個『詞彙特徵值』相同或相似時，便可將該語音訊號片段辨識為該『詞彙模型』所對應的『詞彙』。反之，如果語者對系統預定的『詞彙』不能確定，因而說出和系統預定不同的詞彙時，其『特徵值』便不能對應任一個模型，自然無法辨識出任何結果。

以美國專利 5,218,668 為例，其方法是在詞彙特徵模型比對 (Pattern Comparison) 時將語音訊號的特徵值，和『系統詞彙〈System Vocabulary〉』以及『非系統詞彙〈Filler〉』作比較，來決定辨識的結果。其問題在於，當語者無法正確的說出系統詞彙時，便會

被辨識成非系統詞彙或是被誤判為某個系統詞彙，即便語者知道該系統詞彙的大部分。

而根據本發明之利用“未知的代稱”(Unknown Code Word)，來增加語音辨識系統的彈性，使得語者在不確定系統『詞彙』的時候，可以採用模糊的語法來進行查詢。而此“未知的代稱”(Unknown Code Word，底下簡稱 UCW)，為一個或多個固定的片語，其功用在於取代一個『詞彙』中不確定的部分。

例如，在使用中文時，經常利用『什麼』來代替不確定的字、詞；如使用『什麼』這個詞當作系統的 UCW 時，當某個使用者打算查詢有關於元朝畫家『趙孟頫』的資料時，卻突然忘記了『頫』這個字個念法，或是不能確定第三個字是什麼。

根據本發明之實施例，語者便可以用『趙孟什麼』這樣配合 UCW 來進行查詢。而辨識系統在得到辨識的結果後，便可利用『趙』以及『孟』這兩個資訊來進行查詢。

請參照圖 2，是說明本發明一較佳實施例之詞彙產生裝置及其方法之電路方塊圖。此詞彙產生裝置 200 包括一斷詞單元(Tokenize Unit)220、一詞彙產生單元(Vocabulary Generation Unit)230 與一“未知的代稱”(Unknown Code Word，UCW)資料庫 240。本發明較佳實施例之詞彙產生裝置及其方法，首先，斷詞單元(Tokenize Unit)220 用以連接到一原始系統

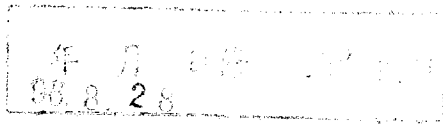
詞彙資料庫 210，並用以接收系統詞彙。此原始系統詞彙資料庫 210 是語音辨識系統所原有的系統詞彙資料庫。而未知的代稱(UCW)資料庫 240 之內容為系統事先定義的詞彙，以及使用者所自行定義之多數個詞彙。

此斷詞單元 220 根據所接收的系統詞彙進行斷詞(Tokenize)操作。主要是將個別之系統詞彙分割為較小的單位之含有意義之單位(Meaningful Unit，底下簡稱“MU”)。而這裡所謂的含有意義之單位可以是字或是詞，依照整個辨識系統的需要而定。

而後，詞彙產生單元(Vocabulary Generation Unit)230 用以接收斷詞單元 220 所產生的含有意義之單位(MU)，並根據“未知的代稱”(Unknown Code Word, UCW)資料庫 240 之內容，將詞彙合併(Vocabulary Generation)，以便將斷詞的結果配合 UCW 結合成新的系統詞彙，並輸出至新系統詞彙資料庫 250。

在一實施例中，產生新的系統詞彙時，可以捨棄部分的 MU 以 UCW 代替，例如，『亡羊補牢』可以產生『亡什麼補牢』以及『什麼補牢』等詞彙。這就是本發明基本的利用“未知的代稱”(UCW)，來增加語音辨識系統的彈性，使得語者在不確定系統『詞彙』的時候，可以採用模糊的語法來進行查詢。

在另外一實施例中，產生新的詞彙時可以捨棄部



分的 MU，例如，『中華民國專利公報』可以產生『專利公報』以及『中華專利公報』等詞彙。也就是說，本發明除了利用“未知的代稱”(UCW)，來增加語音辨識系統的彈性使得語者在不確定系統『詞彙』的時候，可以採用模糊的語法來進行查詢。也可重組整個原始的系統詞彙，產生重組後新的系統詞彙。

在本發明之一實施例中，產生新的詞彙時可以直接插入 UCW，例如『專利公報』可以產生『專利什麼公報』以及『什麼專利公報』等詞彙。

而在本發明之另一實施例中，亦可在產生新的詞彙時將 MU 調換順序，例如『中華民國專利公報』可以產生『專利什麼中華民國公報』。

本發明之利用“未知的代稱”(UCW)，使得語者在不確定系統『詞彙』的時候，可以採用模糊的語法來進行查詢，在一實施例中，可以在產生新的詞彙時可以加入一個以上的 UCW，例如『中華民國專利公報』可以產生『中華民國什麼專利什麼公報』。此可更有效地增加語音辨識系統的彈性。也可同時採取以上的方法，例如，『中華民國專利公報』可以產生『專利什麼公報』以及『中華專利什麼公報』等詞彙。

本發明所提出的詞彙產生裝置及其方法，主要是用以產生新的系統詞彙資料庫。而底下將說明如何將本發明用以產生新系統詞彙資料庫後，如何具體運用在實際的語音辨識系統中。以一個電視節目的語音系統

為例，台視的新聞相關節目一共包括了：『早安您好六點半新聞』、『早安您好七點新聞』、『台視台語新聞』、『台視 1200 午間新聞』、『午安您好-台視新聞』和『台視晚間新聞』六個不同時段的新聞節目。對於一般的觀眾而言，如果需要記住節目完整的名稱，才能使用語音的功能實在是太過不方便。如果採用本專利的技術來對這六個節目名稱進行處理後，可以產生新的系統詞彙，如：『早安什麼新聞』，『早安您好什麼新聞』，『台視什麼新聞』，『台視新聞』，『台視午間什麼新聞』，『什麼台語新聞』等等數個系統詞彙。而使用者便可使用這些新增的新系統詞彙來進行查詢，例如『早安什麼新聞』便可對應到『早安您好六點半新聞』，『早安您好七點新聞』兩個節目，之後系統便可列出這兩個節目，再由使用者來進行選擇。而『什麼台語新聞』則可以對應到『台視台語新聞』，這個節目。依此，經過本專利的處理後，系統詞彙的數量將會擴大，產生的系統詞彙有的對應到單一個節目，也有的對應到複數的節目，而使用這一類系統詞彙的好處在於使用者不需要記憶完整的系統詞彙，僅需記憶系統詞彙中重要的部分，便可使用系統的功能。因此，運用本發明所提出的詞彙產生裝置及其方法之語音辨識系統，可增加語音辨識系統的彈性，使得使用者在不能準確的說出系統可辨識的語音命令時，可以使用特別定義的一種“未知的代稱”(Unknown Code

Word)來代替命令中不能確定的部分，以提升使用者之使用方便性。

雖然本發明已以較佳實施例揭露如上，然其並非用以限定本發明，任何熟習此技藝者，在不脫離本發明之精神和範圍內，當可作些許之更動與潤飾，因此本發明之保護範圍當視後附之申請專利範圍所界定者為準。

【圖式簡單說明】

圖 1 是說明語音辨識的主要流程。

圖 2 是說明本發明一較佳實施例之詞彙產生裝置及其方法之電路方塊圖。

【主要元件符號說明】

- 100 語音辨識系統
- 110 語音特徵值擷取單元
- 120 詞彙特徵模型比對單元
- 130 詞彙資料庫
- 140 辨識決定單元
- 200 詞彙產生裝置
- 210 原始系統詞彙資料庫
- 220 斷詞單元(Tokenize Unit)
- 230 詞彙產生單元(Vocabulary Generation Unit)
- 240 UCW 資料庫
- 250 新系統詞彙資料庫

五、中文發明摘要：

一種詞彙產生方法、裝置及運用上述詞彙產生裝置之語音辨識系統。此詞彙產生方法可產生新的系統詞彙，以增加語音辨識系統的彈性，使得使用者在不能準確的說出系統命令時，可以使用本發明特別定義的一種“未知的代稱”(Unknown Code Word)來代替命令中不能確定的部分。

六、英文發明摘要：

A vocabulary generating apparatus and method thereof and a speech recognition system with the vocabulary generating apparatus are provided. The vocabulary generating method can generate new system recognition vocabularies for improving the flexibility of speech recognition. When the user uses the speech recognition system for operation and forgets some vocabularies for commands, the user can use unknown code word (UCW) to replace the forgotten vocabularies.

十、申請專利範圍：

1、一種詞彙產生裝置，包括

一斷詞單元，用以接收多數個原始系統詞彙，並進行斷詞(Tokenize)操作，並分割每個該原始系統詞彙為對應之多數含有意義之單位，其中該些原始系統詞彙是來自一語音辨識系統；

一未知的代稱資料庫，用以儲存多數個未知的代稱；以及

一詞彙產生單元，連接到該斷詞單元與該未知的代稱資料庫，用以接收該些含有意義之單位以及該些未知的代稱，並進行詞彙合併操作，以便將該斷詞單元所產生該些含有意義之單位結合該些未知的代稱而合成多數個新系統詞彙。

2、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該斷詞單元根據每個該原始系統詞彙產生對應之該些含有意義之單位為多數個字。

3、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該斷詞單元根據每個該原始系統詞彙產生對應之該些含有意義之單位為多數個詞。

4、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該些原始系統詞彙是來自該語音辨識系統所原有的系統詞彙資料庫內之多數個系統詞彙。

5、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該些未知的代稱詞彙是根據使用者所自行

定義之多數個詞彙。

6、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是捨棄部分之含有意義之單位而以未知的代稱所代替。

7、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位之部分捨棄，並加入該未知的代稱所代替。

8、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入該未知的代稱。

9、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入多數個該些未知的代稱。

10、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該

些新系統詞彙是將含有意義之單位先行重組後直接加入該些未知的代稱而產生該些新系統詞彙。

11、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是省略部分有意義之單位後再進行重組以產生新的系統詞彙。

12、如申請專利範圍第 1 項所述之一種詞彙產生裝置，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是將該些未知的代稱加入以產生新的系統詞彙。

13、一種詞彙產生方法，適用於一語音辨識系統，該語音辨識系統具有一原始系統詞彙資料庫，該方法包括：

將該原始系統詞彙資料庫內的多數個原始系統詞彙進行斷詞(Tokenize)操作，並分割每個該原始系統詞彙為對應之多數含有意義之單位；

將該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以便將該斷詞單元所產生該些含有意義之單位結合該些未知的代稱而合成多數個新系統詞彙，以提供該語音辨識系統用以進行語音辨識。

14、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據每個該原始系統詞彙產生對應之該些含有意義之單位為多數個字。

15、如申請專利範圍第 13 項所述之一種詞彙產

生方法，其中根據每個該原始系統詞彙產生對應之該些含有意義之單位為多數個詞。

16、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是捨棄部分之含有意義之單位而以未知的代稱所代替。

17、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位之部分捨棄，並加入該未知的代稱所代替。

18、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入該未知的代稱。

19、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入多數個該些未知的代稱。

20、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位先行重組後直接加入該些未知的代稱而產生該些新系統詞彙。

21、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是省略部分含有意義之單位後再進行重組以產生新的系統詞彙。

22、如申請專利範圍第 13 項所述之一種詞彙產生方法，其中根據該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是加入該些未知的代稱以產生新的系統詞彙。

23、一種具有詞彙產生裝置之語音辨識系統，包括：

一原始系統詞彙資料庫，用以儲存多數個原始系統詞彙；

一未知的代稱資料庫，用以儲存多數個未知的代稱；

一斷詞單元，用以接收該些原始系統詞彙，並進行斷詞(Tokenize)操作，並分割每個該原始系統詞彙為對應之多數含有意義之單位；以及

一詞彙產生單元，連接到該斷詞單元與該未知的代稱資料庫，用以接收該些含有意義之單位以及該些未知的代稱，並進行詞彙合併操作，以便將該斷詞單元所產生該些含有意義之單位結合該些未知的代稱而合成多數個新系統詞彙。

24、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位

以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是捨棄部分之含有意義之單位而以未知的代稱所代替。

25、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位之部分捨棄，並加入該未知的代稱所代替。

26、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入該未知的代稱。

27、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位直接加入多數個該些未知的代稱。

28、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，以產生該些新系統詞彙是將含有意義之單位先行重組後直接加入該些未知的代稱而產生該些新系統詞彙。

29、如申請專利範圍第 23 項所述之語音辨識系

統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是將部分之含有意義之單位省略後再進行重組以產生新的系統詞彙。

30、如申請專利範圍第 23 項所述之語音辨識系統，其中該詞彙產生單元將接收該些含有意義之單位以及該些未知的代稱進行詞彙合併操作，是將該些未知的代稱加入以產生新的系統詞彙。

96-8-28

七、指定代表圖：

(一)本案指定代表圖為：第 (2) 圖。

(二)本代表圖之元件符號簡單說明：

- 200 詞彙產生裝置
- 210 原始系統詞彙資料庫
- 220 斷詞單元(Tokenize Unit)
- 230 詞彙產生單元(Vocabulary Generation Unit)
- 240 UCW 資料庫
- 250 新系統詞彙資料庫

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無。