

公告本

發明專利說明書 I223165

(填寫本書件時請先行詳閱申請書後之申請須知，作※記號部分請勿填寫)

※ 申請案號： 91121867 ※IPC 分類： G06F 17/21
※ 申請日期： 91-9-24

壹、發明名稱

(中文) 日文虛擬字典
(英文) JAPANESE VIRTUAL DICTIONARY

貳、發明人 (共 2 人)

發明人 1 (如發明人超過一人，請填說明書發明人續頁)

姓名：(中文) 紀田 康雄
(英文) YASUO KIDA

住居所地址：(中文) 美國加州帕羅艾托市哈克大道 1025 號
(英文) 1025 HARKER AVENUE, PALO ALTO, CALIFORNIA
94301, U.S.A.

國籍：(中文) 日本 (英文) JAPAN

參、申請人 (共 1 人)

申請人 1 (如申請人超過一人，請填說明書申請人續頁)

姓名或名稱：(中文) 美商蘋果電腦公司
(英文) APPLE COMPUTER, INC.

住居所或營業所地址：(中文) 美國加州庫柏提諾市伊芬尼羅佩街 1 號
(英文) 1 INFINITE LOOP, CUPERTINO, CALIFORNIA
95014, U.S.A.

國籍：(中文) 美國 (英文) U.S.A.

代表人：(中文) 理察 J. 路頓二世
(英文) RICHARD J. LUTTON, JR.

發明人 2

姓名：(中文) 原 啟介

(英文) KEISUKE HARA

住居所地址：(中文) 日本國川崎縣川崎市宮前區鷺沼 1-18-11-404

(英文) 1-18-11-404 SAGINUMA, MIYAMAE-KU, KAWASAKI-SI,

KAWASAKI, JAPAN 216-0004

國籍：(中文) 日本

(英文) JAPAN

捌、聲明事項

☐ 本案係符合專利法第二十條第一項☐第一款但書或☐第二款但書規定之期間，其日期為：_____

☒ 本案已向下列國家（地區）申請專利，申請日期及案號資料如下：

【格式請依：申請國家（地區）；申請日期；申請案號 順序註記】

1. 美國；2001 年 09 月 25 日；09/965,333

2. _____

3. _____

☒ 主張專利法第二十四條第一項優先權：

【格式請依：受理國家（地區）；日期；案號 順序註記】

1. 美國；2001 年 09 月 25 日；09/965,333

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

☐ 主張專利法第二十五條之一第一項優先權：

【格式請依：申請日；申請案號 順序註記】

1. _____

2. _____

3. _____

☐ 主張專利法第二十六條微生物：

☐ 國內微生物 【格式請依：寄存機構；日期；號碼 順序註記】

1. _____

2. _____

3. _____

☐ 國外微生物 【格式請依：寄存國名；機構；日期；號碼 順序註記】

1. _____

2. _____

3. _____

☐ 熟習該項技術者易於獲得，不須寄存。

(1)

玖、發明說明

(發明說明應敘明：發明所屬之技術領域、先前技術、內容、實施方式及圖式簡單說明)-

發明領域

本發明一般係關於電子文字輸入的領域；更明確地說，其係關於一種方法，用以輸入日文平假名字元，並且利用平假名、片假名及漢字字元組合將其翻譯成日文單詞。

發明背景

日文的書寫語言包括三種分離的字元字串。平假名及片假名字元組(合稱為「假名」)代表的是發音音節的簡單日文字元。平假名字元(其特徵為草寫型體)通常係用以書寫日本原有的單詞。片假名字元(其特徵為型體較為工整)通常係用以書寫從外來文化假借過來的單詞，或是為了強調及發音的效果。日文中的第三種字元組則是漢字。漢字是從中國語言假借而來的複雜日文字元。在日文語言中共有超過9000個漢字。大約有4000個漢字可半正式地使用，而如果要閱讀一份報紙或是在日本閒逛的話，通常需要認識2000個漢字。日文書寫語言的複雜性對於想要以電腦、文字處理器及其它電子元件作有效的文字輸入來說，極具挑戰性。

圖1A顯示的係日文平假名及片假名字元實例。平假名151及片假名152字元組各含有46個基本字元。兩組假名具有相同的發音及構造規則，差別僅在於字元的形狀，以強調不同的單詞用法。某些基本的假名字元會使用於特定的組合中並且配合特殊的符號(所謂的「濁音符號(nigori)」及「半濁音符號(maru)」)，形成基本音節的濁音及送氣音變化，因此，可形成約120種日文發音的完整字元組。如果日文鍵

盤包含基本音節所有濁音及送氣音變化的個別按鍵的話，那麼至少需要80個字元按鍵。如此龐大的按鍵數量會使得鍵盤的按鍵過多，不易使用。如果分開包含濁音符號及半濁音符號按鍵的話，便可將字元按鍵的數量降低成57個按鍵。不過，為產生基本字元的濁音及送氣音，使用者必須對單一個字元進行兩次或更多次的敲擊。

在電腦及類似元件中輸入日文字的常見方法，通常是要使用適合日文使用的標準日文字元鍵盤或羅馬字元鍵盤。一般的假名鍵盤僅具有一種代表可以直接由該鍵盤輸入的假名組(通常是平假名)的按鍵。慣用的方法是以含有平假名按鍵的鍵盤的平假名文字當作輸入，並且使用所謂的假名-漢字轉換方式將其轉換成日文。一般的日文都係以平假名、片假名及漢字字元來表示，例如句子150，其英文句意為「在聖荷西看電影」。原文150包括片假名字元154，其相當於外來字「聖荷西」；平假名字元155，其一般是當作助詞；以及漢字字元組153。

圖1B所示的係將平假名內容轉換成日文內容的慣用方法。參考圖1，在101中，可透過鍵盤輸入日文的平假名字元。查詢資料庫(例如字典)將該些平假名字元轉換102成日文內容。接著，使用者必須檢查103及核對104是否轉換正確。如果轉換不正確的話(例如，該字典不含此項轉換)，使用者必須以手動的方式強迫該系統轉換該平假名內容。一般的使用者互動包含選取105轉換不正確的平假名內容部分，明確地指示106該系統轉換此部份。接著，該系統便

會呈現 107 一份包括所有可能選項的候選清單。使用者通常會核對 109 轉換是否正確。如果轉換正確的話，那麼使用者便會選取 108 其中一個選項，作為最佳的輸出，並且插入該正確的結果，形成最終的輸出內容。如果轉換不正確的話，使用者便必須重新選取不同的輸入部分，並且試著以手動的方式再度轉換該重新選取的部分。

在大部分的情形中，慣用的方法中，音譯(從平假名或片假名直接轉換)通常無法產生正確的結果，因為通常使用者都會選取(例如圖 1B 所示的方法)含有該片假名單詞的小節(而非該片假名單詞本身)以及以平假名書寫的一個或多個語末助詞作為最終的形式。一般的音譯法亦會將所有的語末助詞轉換成不正確的片假名形式。

另一種慣用的方法則是對該字串的前面子字串進行音譯，產生替代的候選項。此方法的優點是，語末助詞一定是在尾部，並且全部都是平假名的形式。此方法會產生許多的候選項，其中包含正確的結果。下面的實例闡述的便是該種慣用的方法(英文的形式)：

輸入：inthehouse

輸出 1：INTHEHOUSE

輸出 2：iNTHEHOUSE

輸出 3：inTHEHOUSE

輸出 4：intHEHOUSE

輸出 5：inthEHOUSE

輸出 6：intheHOUSE-(正確結果)

輸出 7：inthehOUSE

輸出 8：inthehoUSE

輸出 9：inthehouSE

輸出 10：inthehouseE

輸出 11：inthehouse

如上述，當使用者選擇欲校正的輸入內容(其可能會混淆最終的選取結果)之後，該慣用的方法便會產生許多候選項，不過其中可能僅包括一個正確選項。另一種慣用的方法則含有一分析器，其能夠辨別後端的助詞。其可從結束位置開始分析，直到該分析器無法找到任何的助詞為止。然而，該慣用的方法必須與使用者進行互動，因而為得到正確的結果，可能會降低其效率。

該慣用的方法的其中一項缺點是，如果在該字典中沒有片假名單詞的話，含有該片假名的轉換結果通常都會失敗。此方法的另一項缺點是，其需要與特定的使用者進行互動方能轉換及選擇最佳的候選項。如果該使用者不熟悉可能的轉換輸出的話，將會花費許多的時間與精力。所以，吾人非常希望有一種更佳的方法，其能夠自動且有效地將日文平假名字元字串轉換成片假名字元字串。

發明概述

本發明所發表的是用以將第一字元字串轉換成第二字元字串的方法及裝置。除了一般的字典之外，本發明包括一虛擬字典，用以根據該第一字元字串產生人造的字元字串。當無法透過一般的字典轉換該第一字元字串時(例如，

該一般的字典不懂該第一字元字串)，本發明便可使用虛擬字典所產生的人造字元字串來轉換該第一字元字串。所以，利用本發明的虛擬字典進行轉換絕對不會失敗。

在本發明的其中一項觀點中，提供一種示範方法，其包括接收平假名輸入；根據該平假名自動決定複數個可能的片假名候選項；分析該複數個可能的片假名候選項，以便將平假名輸入轉換成片假名字元；選擇其中一個片假名候選項；以及輸出包含該個片假名候選項與(在某些情形中)漢字字元的轉換內容。

在本發明的另一項觀點中，提供一種示範方法，其包括接收具有原始字元字串之第一字元字串；將該第一字元字串分割成複數個子字串；透過字典將該複數個子字串轉換成第二字元字串；產生與該複數個子字串對應的第三字元字串；分析第二及第三字元字串；根據分析結果，由第二及第三字元字串建立第四字元字串；根據第四字元字串產生一份候選清單；從該候選清單中選出目標字元字串並且將其輸出。

在其中一種特殊的示範具體實施例中，該方法包括，如果第二字元字串包含對應於第一字元字串的字元字串的話，便由第二字元字串建立第四字元字串；如果第二字元字串不包含對應於第一字元字串的字元字串的話，便由第三字元字串建立第四字元字串。在另一具體實施例中，該方法包括，檢查轉換的輸出結果，判斷轉換是否正確，如果轉換不正確的話便提供替代字元字串候選清單，並且從

該候選清單中選擇一字元字串作為最終的輸出。在進一步的具體實施例中，該方法包括提供一人造的目標字元字串，並且根據該人造的字元字串更新資料庫。

本發明包括執行該些方法的裝置，以及機器可讀取的媒體，當在一資料處理系統中執行的時候，其便可讓該系統執行該些方法。從隨附的圖式及後面的詳細說明中將會非常清楚本發明的其它特點。

圖式簡單說明

本發明將透過隨附圖式進行說明(但不受限於此)，其中，相同的參考符號代表相同的元件。

圖1A所示的係日文字元實例，包括平假名、片假名及漢字字元。

圖1B所示的係用以將平假名內容轉換成日文內容的慣用方法。

圖2所示的係可與本發明共同使用的電腦系統。

圖3所示的係本發明的假名-漢字轉換系統之一具體實施例。

圖4所示的係本發明一具體實施例所使用的片假名字元組之成本值計算實例。

圖5所示的係本發明中具有使用者互動的假名-漢字轉換系統之另一具體實施例。

圖6A所示的係根據本發明從平假名字元組轉換成片假名字元組之轉換過程的一具體實施例。

圖6B所示的係相較於慣用方法過程的本發明實例之示意

圖。

圖 7 所示的係本發明將平假名字元轉換成片假名字元的方法。

圖 8 所示的係根據本發明從平假名字元組轉換成片假名字元組之轉換過程的另一具體實施例。

圖 9A 及 9B 所示的係本發明將平假名字元轉換成片假名字元的另一種方法。

詳細說明

下面的說明及圖式僅係闡述本發明，而非限制本發明。其中將敘述各種特定的細節，以便能夠對本發明有澈底的瞭解。不過，在某些情形中，熟知的或慣用的細節則不會加以敘述，以免對本發明的細節造成不必要的混淆。

日文的書寫方式包含漢字(源自於中文字元)及兩組發音的假名符號，即平假名與片假名。單一個漢字字元可能包含一個符號或數個符號，並且本身可能就代表整個單詞或欲表達的對象。假名與漢字不同的地分在於，除非與其它的假名或漢字結合形成單詞，否則本身並不具意義。平假名及片假名各包含 46 個符號。假名字元的組合及變化構成日語中的所有發聲基礎。所有的日文內容都能夠以平假名或片假名的方式來書寫。不過，因為日文中的單詞之間並沒有間隔，所以，當某段句子中的單詞僅由平假名或片假名構成的時候，閱讀起來會相當地不方便。所以，大部分的日文內容都會包括平假名、片假名及漢字字元。一般來說，漢字字元都是當作名詞、形容詞或動詞；而平假名及

片假名則是當作助詞(例如「 of」、「 at」等)。

當電腦化的文字處理機經過大幅改良之後，便能夠經由文字處理軟體實現日文的文字處理。將所有的平假名、片假名及漢字字元(假名-漢字)都包含在一個鍵盤中的做法不大實際，因此，一般的方式都是將日文字元以平假名的方式輸入。所以，大部分的重點便會放在提供一種將平假名字元轉換成片假名字元的改良方法上。本發明便提供一種獨特的方法，其係根據平假名字元與片假名字元之間預設的關係，自動將平假名字元轉換成片假名字元。該些方法通常都可由電腦中所執行的軟體來實施。

圖2所示的係可與本發明共同使用的典型電腦系統。請注意，雖然圖2顯示各種電腦系統的組件，不過並不代表連接該些組件的任何特殊架構或方式，因為此部份細節並非本發明的重點。亦可發現到，具有較少或較多組件的網路電腦及其它的資料處理系統(例如個人數位助理)亦可能與本發明共同使用。舉例來說，圖2的電腦系統可能是Apple Macintosh電腦或個人數位助理(PDA)。

如圖2所示，電腦系統200(其為一種資料處理系統形式)包括一條匯流排202，其耦合至微處理器203、ROM 207、揮發性RAM 205及非揮發性記憶體206。如圖2範例所示，微處理器203(其可能是Motorola Inc.所售的G3或G4微處理器或是IBM所售的微處理器)會耦合至快取記憶體204。匯流排202會將該些不同的組件相連在一起，亦會將該些組件203、207、205及206連接至顯示控制器及顯示元件208及如輸入/

輸出(I/O)元件般的週邊元件，例如滑鼠、鍵盤、數據機、網路介面、印表機及本技藝熟知的其它元件。一般來說，輸入/輸出元件210都係透過輸入/輸出控制器209與該系統耦合。揮發性RAM 205通常是實現成動態RAM (DRAM)，其需要持續的電源，方能補充或維持該記憶體中的資料。非揮發性記憶體206則通常是磁性硬碟機、磁性光碟機、光碟機、DVD RAM或其它類型的記憶體系統，即使拔除該系統的電源，其亦可維持資料。一般來說，該非揮發性記憶體亦可能是隨機存取記憶體，不過卻非絕對。雖然圖2中所顯示的非揮發性記憶體是直接耦合至該資料處理系統中其餘組件的區域性元件，不過可發現到，本發明亦可能運用與該系統相隔很遠的非揮發性記憶體，例如透過網路介面(例如數據機或乙太網路介面)耦合至該資料處理系統的網路儲存元件。匯流排202可能包括一條或多條相互連接的匯流排，其係透過各種橋接器、控制器及/或本技藝所熟知的轉接器進行連接。在其中一種具體實施例中，I/O控制器209包括USB(通用序列匯排流)轉接器，用以控制USB週邊。

圖3所示的係本發明一具體實施例所使用的系統。參考圖3，系統300通常包括一輸入單元301、一輸入方法UI及系統介面302、一形態分析引擎(MAE) 303、一字典管理模組(DMM) 305及一輸出單元308。輸入單元301可能是如圖2的I/O元件210之類的鍵盤。該輸入單元亦可能是觸控墊，例如個人數位助理(PDA)。該輸入單元亦可能是可從應用程式接收輸入的一組應用程式介面(API)。當然還有其它類型的輸

入。輸入單元 301 可接受日文字元輸入(例如,日文平假名字元)。該些平假名字元會被傳送至輸入方法及系統介面 302,然後再被傳送至 MAE 303。接著,MAE 303 便會透過 DMM 305 存取資料庫,例如一般字典 307 及虛擬字典 306。一般字典 307 可能包括對應於平假名單詞的最熟悉的日文單詞。該一般字典 307 可能會儲存在隨機存取記憶體(RAM)之中,例如揮發性的 RAM 205;或是可能會儲存在硬碟之中,例如非揮發性的記憶體 206。在其中一具體實施例中,該一般字典 307 則可能會透過網路儲存在遠端的儲存位置中(例如網路儲存體)。值得注意的是,本發明可能實現於網路計算環境中,所以,該些一般字典可能會儲存在伺服器中,而在客端電腦執行的應用程式則可透過網路中的網路介面存取該些一般字典。在多部客端電腦執行的多個應用程式可同時存取該些一般字典,並且透過該網路分享該些一般字典的資訊。雖然圖中的一般字典 307 顯示成單個字典,不過可發現到,該一般字典 307 亦可能包括多個字典或資料庫。在另一具體實施例中,該些一般字典 307 可能包括多個對照表。虛擬字典 306 可能會將每個單平假名字元直接轉換成單片假名字元。該虛擬字典可能含有一個對照表,用以查出每個平假名字元所對應的片假名字元。DMM 305 負責管理所有的字典,包括字典 306、307。當接收到 MAE 303 的要求時,DMM 305 亦負責更新該些字典的任何資訊。在其中一具體實施例中,該 DMM 305 亦會管理另一個儲存著所有規則或策略的資料庫 304。

虛擬字典 306 可能會將平假名字元直接翻譯成片假名字元。該虛擬字典 306 可能會回傳具有不同語音部分的所有多個單詞。在其中一具體實施例中，該虛擬字典會回傳三個語音部分。其為名詞，以及可當作動詞及形容詞的名詞。值得注意的是，一但從該虛擬字典回傳之後，由該虛擬字典所產生的人造片假名單詞與一般的單詞並無任何不同。

在另一具體實施例中，可能會將該字典資料庫分割成兩個或多個字典。其中一個是包含一般單詞的一般字典。另外的字典則是特殊字典(例如所謂的虛擬字典)。該特殊字典可能包含所有可能的片假名字元，包括處理期間所產生的人造片假名字元。該片假名係從平假名輸入直接直譯而成的。該虛擬字典可能會回傳具有不同語音部分的多個單詞。每個單詞都具有其優先值。此優先值可能是由該虛擬字典所分配的。舉例來說，在字串「A-Ka-Ma-I」的實現情形中，該字典可能會回傳三個具有不同語音部分的輸出，名詞，與動詞及形容詞有關聯的名詞，如下所示：

Ka-Ma-I POS：名詞優先值：100

Ka-Ma-I POS：可當作動詞的名詞優先值：100

Ka-Ma-I POS：形容詞優先值：100

亦可以其它的方式實施例現。

三個單詞可視為一筆記錄，或可將其視為三筆不同的記錄。從該字典回傳的所有單詞的優先值可能都一樣。亦可從片假名及/或語音部分計算出其優先值。在其中一具體實施例中，其係由該單詞的長度決定其優先值。在另一具體

實施例中，則可依照片假名的雙重節及三重節統計值決定其優先值，並且依照語音部分進行調整。一般來說，在一般字典中，優先值都會設定成低於所有的(或大部分的)一般單詞，以避免有適當的一般單詞可用的時候，卻以人造的片假名單詞作為最可能的轉換。

語音部分則是定義某個語音部分的單詞會在多久之後(或是其出現的容易度)出現在某個語音部分的其它單詞後面。其可能僅為是/否值。受到實現方式的限制，在許多情形中，一個單詞會具有兩個語音部分。一個作為右連接詞，另一個則作為左連接詞。同樣地，在許多情形中，其不僅可用以決定下一個或前一個單詞，亦可與前一個單詞前面或下一個單詞後面的單詞連接。

參考圖3，MAE 303會送出要求給DMM 305，用以轉換所輸入的平假名單詞。該DMM 305會在一般字典307中搜尋對應的日文單詞。同時，該MAE 303會送出要求給DMM 305，從虛擬字典306中擷取出所有可能的片假名字元組合。通常，如果一般字典307含有此類直接翻譯的話，MAE 303便會從中選擇單詞。否則，MAE 303便會選擇虛擬字典306所創造的人造片假名單詞。

MAE 303亦會從資料庫304中呼叫一組規則，並且應用該組規則分析所有可能的組合。含有該些規則的資料庫304可能是一分離的資料庫，或是與字典306或307相同的資料庫。每種可能的組合都與使用頻率有關。使用頻率代表的是前面使用該些字元的頻率。該字典可能亦包括每個字元組(例

如，名詞、形容詞及動詞等)之間的連接關係。該組規則可能包括使用頻率及連接關係的資訊。該 MAE 303 會應用該些規則，根據該組規則從字典 306 的可能組合中建立一份可能的候選名單或清單。在其中一具體實施例中，該組規則可能包括用以建立該份候選清單的語義或文法規則。舉例來說，單詞「hot」可能表示高溫或辛辣的食物。當該單詞「hot」與單詞「summer」結合之後，例如單詞「hot summer」，那麼單詞「hot」便比較可能表示「高溫」，而非「辛辣的」。該 MAE 303 可能會根據該組規則計算該些候選單詞的成本值。最終的候選單詞可能是由該份候選清單中成本值最小者所構成。

圖 4 所示的實例中，有兩個代表單詞「San Jose」的候選單詞，每個都包含一使用頻率。第一種選擇包含字元 401、402，而第二種選擇則包含字元組 404。字元 403 則是助詞。字元 401 的使用頻率為 f_1 ，字元 402 的使用頻率為 f_2 。助詞字元 403 的使用頻率為 f_3 。此外，字元 401 及 402 之間的連接為 c_1 ，而字元 402 及 403 之間的連接為 c_2 。因此，第一種選擇的成本值可能是：

$$\text{成本值 } 1 = f_1 + f_2 + f_3 + c_1 + c_2$$

同樣地，第一種選擇的成本值可能是：

$$\text{成本值 } 2 = f_a + f_3 + c_a$$

在其中一具體實施例中，成本值可能包括語義或文法係數。估算單元 303 會估算兩種選擇的成本值，並且選擇成本值最小的(此例中為成本值 2)一個作為該轉換的最終輸出。

不過，雖然估算單元會依照最小成本值選擇最終的輸出，而且在大部分的情形中所選擇的輸出都是正確的，僅有極少數的情形中正確的輸出並非具有最小成本值。在此情況下，本發明提供使用者進行互動的機會。圖5所示的係本發明之另一具體實施例。參考圖5，系統300提供一使用者互動單元309，讓使用者能夠檢查MAE 303所產生的輸出，並且判斷該輸出是否正確。如果使用者確定該輸出並不正確，MAE 303便會透過DMM 305從資料庫(例如虛擬字典306)中擷取出該候選清單，並將該候選清單顯示在使用者介面中。在其中一具體實施例中，該使用者介面可能是一彈出式視窗。然後，使用者便能夠從該候選清單中選出最佳的選擇(例如最終的選擇)作為輸出。在進一步的具體實施例中，可能會透過應用程式介面(API)將該輸出傳送給應用程式，讓該應用程式選出最終的選擇。

在另一具體實施例中，如果該候選清單仍未包含使用者所需要的正確輸出的話，本發明便進一步提供可讓使用者直接以手動方式輸入最終輸出的構件，並強迫該系統將該些平假名字元轉換成片假名字元。接著，該系統便會更新其資料庫(例如虛擬字典306或一般字典307)，以便將使用者所輸入的最終輸出片假名單詞納入，供以後參考使用。在進一步的具體實施例中，實際上，該使用者可能會修改用於轉換的規則，並將該些使用者指定的規則儲存在資料庫304中。

圖6A所示的係本發明一具體實施例之方塊圖。日文平假

名字元字串 601 (其意義為「在 San Jose 看電影」) 會被輸入至該系統。形態分析引擎 (MAE) 604 會比對資料庫 (如字典 307)，搜尋對應的日文單詞。系統會透過使用者介面 616 將 602 部分傳送給形態分析引擎 (MAE) 604。MAE 604 會將輸入分割成複數個子字串，並且與字典管理模組 (DMM) 608 進行溝通，並且比對字典 606 以便直接翻譯每個子字串。同時，DMM 會指示虛擬字典 607 創造與每個子字串對應的所有可能的片假名單詞。因此，便可利用一般字典 606 中的一般日文單詞及虛擬字典 607 人為創造的片假名單詞形成一份單詞名單 605。在其中一具體實施例中，該些日文字元字串 605 中，每個都有一使用頻率值以及每個字元組之間的連接關係資訊。在另一具體實施例中，該些日文字元字串 605 中，每個都有一優先值。一般來說，人造的片假名單詞的優先值會低於一般字典 606 中的一般單詞，以免混淆。也就是說，該系統在挑選人造的片假名單詞之前，會先從一般字典挑選一般的單詞。除非一般字典 606 中找不到對應的一般單詞，該系統才會使用人造的單詞。優先資訊亦可儲存於字典 606 中。接著，MAE 604 便會估算及分析字元字串 605 且應用資料庫 607 中的規則。雖然圖中的資料庫 607 及字典 606 是分離的資料庫，不過，亦可將兩者結合成單一個資料庫。MAE 604 會根據該組規則從字元字串 605 中建立另一組字元字串 610。單詞 610 可視為一份候選清單，成本值最小的單詞其優先值越高 (例如單詞 611)，成本值較高的單詞其優先值越低 (例如單詞 612)。亦可能還有其它的優先值技術。依照

此份候選清單，MAE 604便可選擇優先值較高的候選單詞作為最終的目標字元字串(例如字元字串613)。接著便可將字元字串613應用於其餘的字元字串中，形成最終的句子614。

圖6B所示的係相較於慣用方法，本發明所使用的方法。參考圖6B，日文平假名字元字串651(其意義為「San Jose」)會經由輸入方法輸入。該輸入方法一般會將該輸入分割成數個子字串652，並且使用字典653盡量將每個子字串652轉換成另一組日文單詞654。字典653通常含有大部分已知的單詞，如單詞663。不過，該些字典卻不知道，單詞「San Jose」(如單詞662)。因此，無法使用該些字典進行轉換，因而不能無法使用單詞662。慣用的方法會對該些單詞654進行分析，應用規則664(例如文法規則)，並且產生一份候選清單660。從該份清單中選出單詞661作為最終的候選單詞，不過該單詞是不正確的。因此，使用者必須以手動方式轉換輸入651，以產生正確的轉換結果。

本發明導入一虛擬片假名字典655。除了使用一般字典進行轉換之外，該虛擬字典655亦會接收子字串652，並且產生一組對應的人造片假名單詞656。結合字典653中的一般單詞654及虛擬字典655所產生的人造片假名單詞656，應用規則組，便可創造一份與該些子字串對應的完整單詞組658。因此，每個子字串都具有其對應的轉換字串，其可能是一般日文單詞(例如單詞663)或人造的片假名單詞。接著，本發明便會根據該組規則657，創造一份候選清單658。從該份候選清單中，可選擇具有最高優先值的單詞作為最終的候選

單詞 659。

圖 7 所示的係本發明一具體實施例的一種方法。參考圖 6A 及 7，該方法從輸入日文平假名字元(例如平假名字元字串 601)開始 701。其會將該平假名字元字串分割 702 成數個子字串，並且透過字典(例如字典 606)將每個子字串轉換 708 成日文單詞。同時，該方法會透過虛擬字典 609 創造出與該輸入相關的所有可能的片假名字元字串 703。利用一般單詞及人造的片假名單詞便可構成一份日文單詞名單 605。接著，便會建立 704 一份候選清單 610，其中，成本值較低的候選單詞其優先值越高，成本值較高的候選單詞其優先值越低。人造的片假名單詞的優先值可能是由該虛擬字典所分配的。接著，該方法便會分析 705 該份候選清單，並且根據分析結果選擇 707 最佳的候選單詞 613 (例如成本值最低者)。接著，便會輸出 708 該最終的候選單詞以形成最終的句子 614。

圖 8 所示的係本發明的另一具體實施例。輸入 601 含有日文平假名字元字串，其中 602 部分(例如「San Jose」)無法直接轉換，而 603 部分則可透過一般字典 606 轉換。接著，該系統便會使用虛擬字典 609，針對 602 部分的每個單一子字元字串，創造出所有可能對應的片假名單詞。接著，型態分析引擎(MAE) 604 會根據一組規則建立另一份候選清單 610。該組規則可能包括字元的使用頻率及該些字元間的連接關係資訊。在其中一具體實施例中，該組規則可能包含語義及文法規則。接著便會針對該份清單中的每個候選單詞，計算其成本值。成本值最小的候選單詞其優先值最高，成本

值最大的候選單詞其優先值最低。如圖8所示，在該份清單的候選單詞之間，候選單詞611的優先值最高。因此，估算單元609便會選擇候選單詞611作為該轉換的最終選擇。然而，在極其少數的情形中，該選擇611可能並不正確，此時便需要進行使用者互動615。在使用者互動期間，使用者會選擇部分的輸入(例如英文意義為「San Jose」的602部分)，並且命令該系統進行轉換。該系統會抽出所有候選單詞的名單，例如候選清單610。在其中一具體實施例中，會透過使用者介面(例如彈出式視窗)顯示該份候選清單。從該份清單中，該使用者會選擇最終輸出616，並且形成最終句子614。根據使用者的選擇，該系統可能會更新其資料庫(例如字典606及虛擬字典609)，使得以後的轉換的成功機率大增。

圖9所示的係本發明另一具體實施例的方法，用以將原始字元字串轉換成目標字元字串。該方法會從使用者介面接收具有該原始字元字串之第一字元字串。其會將該第一字元字串分割成數個子字串。然後透過字典將該些子字串轉換成第二字元字串。同時，該方法會透過虛擬字典創造出對應該些子字串的第三字元字串。接著，便會分析第二及第三字元字串，並且根據分析結果建立第四字元字串。接著，便根據優先資訊產生一份候選清單，然後從該候選清單中選出優先值最高的最終候選單詞。

參考圖9，可透過使用者介面(例如鍵盤)接收日文平假名字元字串。在其中一具體實施例中，該使用者介面可能包

括掌上型事務機的觸控墊，或其它的輸入元件。在進一步的具體實施例中，可能會透過應用程式介面(API)從應用程式接收該日文平假名字元字串。該平假名字元字串會被分割902成數個子字串。然後，型態分析引擎(MAE)便會與字典管理模組(DMM)進行溝通，並且透過一般字典將每個子字串轉換903成對應的日文單詞。同時，MAE亦會指示DMM透過虛擬字典創造904與該些子字串對應的所有可能的片假名單詞。接著，該系統便會從該些所有可能的單詞中(包括一般字典中的日文單詞及虛擬字典所產生的人造片假名單詞)建立905可能的候選單詞，並且構成一份候選清單。來自該虛擬字典的可能的片假名單詞選擇可能包括部分語音資訊。該系統可能會使用一組規則來建立該些候選單詞。在其中一具體實施例中，該組規則可能包括每個片假名字元組的使用頻率值以及每種選擇之間的連接關係。在另一具體實施例中，該組規則可能包括單詞的語義或文法規則。此資訊可能會儲存在儲存著所有可能的片假名字元組的資料庫中。在另一具體實施例中，該些規則則可能儲存在一分離的資料庫中。接著，該系統便會從該資料庫中擷取906使用頻率及連接關係，並且應用907語義及文法規則進行分析。根據此資訊，該系統便可計算908所有候選單詞的成本值。接著，便可選取成本值最低的候選單詞作為最終的目標字元組。然後便可將該最終的目標字元組顯示在顯示元件的使用者介面中。

在本發明的另一項觀點中，使用者可能會檢查910假名-

漢字元字串轉換引擎所提供的結果，以核對911是否轉換正確。如果使用者接受此結果的話，便完成轉換。但是，如果轉換不正確的話，使用者便可選取912該輸入部分(例如原來的平假名輸入)，並且明確地指示該系統轉換此部份。接著，該系統便會以候選清單的形式提供所有可能的日文單詞組合(包括人造的片假名單詞)。然後，使用者便可從此份候選清單中進行擷取913，並且顯示在使用者介面中。在其中一具體實施例中，該使用者介面是一彈出式視窗。然後，該使用者便能夠核對該份候選清單是否包含正確的轉換。如果該份候選清單包含正確的轉換的話，該使用者便可從該候選清單中選出915最佳的候選單詞。接著，該系統便會就使用者的選擇方面，更新916其參數(例如使用頻率、連接關係等)資料庫(例如知識庫)。然後便可將該最終選擇輸出917給應用程式。在其中一具體實施例中，如果該份候選清單未包含正確結果的話，該使用者便可能會透過使用者介面，以手動的方式建立918且創造正確結果。當該使用者創造出人造的轉換之後，該系統便會將此結果儲存919在其資料庫中，供以後參考使用。

在前述的詳細說明中，已經參考本發明特定的示範具體實施例對其作說明確地說，。顯而易見的是，在不脫離後面申請專利範圍所提出的本發明廣義精神及範疇下，可對其進行各種修正。因此，所有的詳細說明及圖式都僅係闡述性質而非加以限制。

圖式代表符號說明

150	句子
151	平假名
152	片假名
153	漢字字元組
154	片假名字元
155	平假名字元
200	電腦系統
202	匯流排
203	微處理器
204	快取記憶體
205	揮發性隨機存取記憶體
206	非揮發性記憶體
207	唯讀記憶體
208	顯示控制器及顯示元件
209	輸入/輸出控制器
210	輸入/輸出元件
300	系統
301	輸入單元
302	輸入方法UI及系統介面
303, 604	形態分析引擎

304, 607	規則資料庫
305, 608	字典管理模組
306, 609	虛擬字典
307	一般字典
308	輸出單元
309	使用者互動單元
401, 402, 403	字元
404	字元組
601, 610, 651	平假名字元字串
601, 602, 603	平假名字元字串601的一部份
605	單詞名單
606, 653	字典
611	高優先值的單詞
612	低優先值的單詞
613	最終的目標字元字串
614	最終的句子
615	使用者互動
616	使用者介面
616	最終的輸出
652	子字串
654, 658	單詞組

(23)

發明說明續頁

655	虛擬片假名字典
656	人造片假名單詞
657, 664	規則
659, 661	最終的候選單詞
658, 660	候選清單
662, 663	單詞

肆、中文發明摘要

本文敘述的是用以將原始字元字串轉換成目標字元字串的方法。在本發明的其中一項觀點中，提供一種示範方法，其包括接收具有原始字元字串之第一字元字串；將該第一字元字串分割成複數個子字串；透過字典將該複數個子字串轉換成第二字元字串；產生與該複數個子字串對應的第三字元字串；分析第二及第三字元字串；根據分析結果，由第二及第三字元字串建立第四字元字串；根據第四字元字串產生一份候選清單；從該候選清單中選出目標字元字串並且將其輸出。本文亦敘述其它的方法及裝置。

伍、英文發明摘要

Methods for converting a source character string to a target character string are described herein. In one aspect of the invention, an exemplary method includes receiving a first character string having the source character string, dividing the first character string into a plurality of sub-strings, converting the plurality of the sub-strings to second character strings through a dictionary, creating third character strings corresponding to the plurality of the sub-strings, analyzing the second and third character strings, constructing fourth character strings from the second and third character strings based on the analysis, creating a candidate list based on the fourth character strings, selecting the target character string from the candidate list and outputting the target character string. Other methods and apparatuses are also described.

陸、(一)、本案指定代表圖為：第 1B 圖

(二)、本代表圖之元件代表符號簡單說明：

柒、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

拾、申請專利範圍

1. 一種以機器方式來實現的方法，用以將原始字元字串轉換成目標字元字串，其包括：
 - 接收具有原始字元字串之第一字元字串；
 - 將該第一字元字串分割成複數個子字串；
 - 透過字典將該複數個子字串轉換成第二字元字串；
 - 創造與該複數個子字串對應的第三字元字串；
 - 分析該第二字元字串及該第三字元字串；
 - 根據分析結果，由該第二及該第三字元字串建立第四字元字串；
 - 根據該第四字元字串創造一份候選清單；
 - 從該候選清單中選出目標字元字串；以及
 - 輸出該目標字元字串。
2. 如申請專利範圍第1項之方法，其中該第一字元字串包括日文平假名字元，並且其中，在該接收之前，該字典係以機器可讀取的形式儲存。
3. 如申請專利範圍第1項之方法，其中該第二字元字串包括日文單詞。
4. 如申請專利範圍第1項之方法，其中該第三字元字串包括片假名字元。
5. 如申請專利範圍第1項之方法，進一步包括：
 - 如果該第二字元字串包含對應於該第一字元字串的字元字串的話，便由該第二字元字串建立該第四字元字串；以及
 - 如果該第二字元字串不包含對應於該第一字元字串的

- 字元字串的話，便由該第三字元字串建立該第四字元字串。
6. 如申請專利範圍第1項之方法，進一步包括根據與該第二字元字串及該第三字元字串相關聯的優先資訊，對該第二字元字串及該第三字元字串進行優先排序。
 7. 如申請專利範圍第6項之方法，該優先資訊包括：
每個該第二字元字串的使用頻率；
字元分節長度；以及
該第二字元字串之間的連接關係資訊。
 8. 如申請專利範圍第6項之方法，進一步包括根據該優先資訊，計算該第二及第三字元字串中每一個的優先值。
 9. 如申請專利範圍第1項之方法，進一步包括，從該字典中擷取出該第二字元字串的優先資訊。
 10. 如申請專利範圍第1項之方法，其中該字典包括數個字典。
 11. 如申請專利範圍第1項之方法，其中，會根據一組規則分析該第二及第三字元字串。
 12. 如申請專利範圍第11項之方法，其中，該組規則包括下面的資訊：該些字元字串的使用頻率以及該些字元字串之間的連接關係。
 13. 如申請專利範圍第12項之方法，其中，該組規則進一步包括下面的資訊：字元分節長度以及語義/文法規則。
 14. 如申請專利範圍第11項之方法，其中，會根據該組規則，由該第二及第三字元字串建立該第四字元字串。

15. 如申請專利範圍第1項之方法，進一步包括，計算每一個該第四字元字串的優先值。
16. 如申請專利範圍第15項之方法，其中會根據該第四字元字串的優先值創造該份候選清單。
17. 如申請專利範圍第16項之方法，其中，會從該份候選清單中選出具最高優先值的候選單詞作為該目標字元字串。
18. 如申請專利範圍第1項之方法，進一步包括：
檢查該目標字元字串的輸出結果，判斷轉換是否正確；
如果該輸出不正確的話，便擷取一份替代字元字串候選清單；以及
從該候選清單中選擇一字元字串作為最終的輸出。
19. 如申請專利範圍第18項之方法，其中一使用者會檢查該目標字元字串的輸出。
20. 如申請專利範圍第18項之方法，進一步包括，將該候選清單顯示在一使用者介面中。
21. 如申請專利範圍第18項之方法，其中一使用者會從該候選清單中選擇一字元字串作為最終的輸出。
22. 如申請專利範圍第18項之方法，進一步包括，根據從該候選清單中所選出的該最終輸出，就該最終輸出字元字串的該優先資訊方面更新一資料庫。
23. 如申請專利範圍第20項之方法，其中該使用者介面係一彈出式視窗。
24. 如申請專利範圍第18項之方法，其中，如果該候選清單

未包含正確的目標字元字串的話，該方法進一步包括：

由一使用者提供該目標字元字串；

以手動的方式將該第一字元字串轉換成該目標字元字串；以及

將該使用者所提供的該目標字元字串儲存於一資料庫中。

25. 如申請專利範圍第24項之方法，其中該目標字元字串係由該使用者透過一使用者介面進行輸入。

26. 如申請專利範圍第11項之方法，其中該組規則係由一使用者透過一使用者介面加以定義。

27. 如申請專利範圍第1項之方法，其中係透過對照表創造該第三字元字串。

28. 如申請專利範圍第1項之方法，其中係透過演算法創造該第三字元字串。

29. 如申請專利範圍第8項之方法，進一步包括分配一優先值給每一個該第三字元字串。

30. 如申請專利範圍第29項之方法，其中該第三字元字串的優先值低於該第二字元字串的優先值。

31. 一種以資料處理系統方式來實現的方法，用以將第一日文字元輸入字串轉換成一第二日文字元字串，該方法包括：

接收一平假名輸入；

根據該平假名輸入，自動決定複數個可能的片假名候選項；

分析該複數個可能的片假名候選項，以便將該平假名輸入轉換成片假名字元；

依照分析的結果，選擇其中一個片假名候選項；以及輸出包含該個片假名候選項的轉換內容。

32. 如申請專利範圍第31項之方法，其中該轉換內容進一步包括漢字字元。

33. 一種用以將原始字元字串轉換成目標字元字串的裝置，其包括：

用以接收具有該原始字元字串之一第一字元字串的構件；

用以將該第一字元字串分割成複數個子字串的構件；

用以透過一字典將該複數個子字串轉換成第二字元字串的構件；

用以創造與該複數個子字串對應的第三字元字串的構件；

用以分析該第二字元字串及該第三字元字串的構件；

用以根據該分析結果，由該第二及第三字元字串建立第四字元字串的構件；

用以根據該第四字元字串創造一份候選清單的構件；

用以從該候選清單中選出該目標字元字串的構件；以及

用以輸出該目標字元字串的構件。

34. 如申請專利範圍第33項之裝置，其中該第一字元字串包括日文平假名字元，並且其中在該接收之前，該字典係

以機器可讀取的形式儲存。

35. 如申請專利範圍第33項之裝置，其中該第二字元字串包括日文單詞。

36. 如申請專利範圍第33項之裝置，其中該第三字元字串包括片假名字元。

37. 如申請專利範圍第33項之裝置，進一步包括：

如果該第二字元字串包含對應於該第一字元字串的字元字串的話，該第二字元字串建立該第四字元字串之構件；以及

如果該第二字元字串不包含對應於第一字元字串的字元字串的話，該第三字元字串建立該第四字元字串之構件。

38. 如申請專利範圍第33項之裝置，進一步包括根據與該第二字元字串及該第三字元字串相關聯的優先資訊，對該第二字元字串及該第三字元字串進行優先排序之構件。

39. 如申請專利範圍第38項之裝置，該優先資訊包括：

每個該第二字元字串的使用頻率；

字元分節長度；以及

該等第二字元字串之間的連接關係資訊。

40. 如申請專利範圍第39項之裝置，進一步包括根據該優先資訊，計算該第二及第三字元字串中每一個的優先值之構件。

41. 如申請專利範圍第33項之裝置，進一步包括，從該字典中擷取出該第二字元字串的優先資訊。

42. 如申請專利範圍第33項之裝置，其中該字典包括數個字典。

43. 如申請專利範圍第33項之裝置，其中針對該第二及第三字元字串的分析係根據一組規則而執行的。

44. 如申請專利範圍第43項之裝置，其中該組規則包括下面的資訊：該些字元字串的使用頻率以及該些字元字串之間的連接關係。

45. 如申請專利範圍第44項之裝置，其中，該組規則進一步包括下面的資訊：字元分節長度以及語義/文法規則。

46. 如申請專利範圍第43項之裝置，其中由該第二及第三字元字串建立該第四字元字串係根據該組規則而執行的。

47. 如申請專利範圍第33項之裝置，進一步包括用以計算該等第四字元字串之每一個的優先值的構件。

48. 如申請專利範圍第47項之裝置，其中該份候選清單係根據該第四字元字串的優先值而創造的。

49. 如申請專利範圍第48項之裝置，其中具最高優先值的候選單詞會從該份候選清單中選出作為該目標字元字串。

50. 如申請專利範圍第33項之裝置，進一步包括：

用以檢查該目標字元字串的輸出結果的構件，以便判斷轉換是否正確；

如果該輸出不正確的話，會擷取一份替代字元字串候選清單之構件；以及

用以從該候選清單中選擇一字元字串作為最終的輸出的構件。

51. 如申請專利範圍第50項之裝置，其中一使用者會檢查該目標字元字串的輸出。
52. 如申請專利範圍第50項之裝置，進一步包括用以將該候選清單顯示在一使用者介面中的構件。
53. 如申請專利範圍第50項之裝置，其中一使用者會從該候選清單中選擇一字元字串作為最終的輸出。
54. 如申請專利範圍第50項之裝置，進一步包括用以根據從該候選清單中所選出的該最終輸出，就該最終輸出字元字串的該優先資訊方面更新資料庫之構件。
55. 如申請專利範圍第52項之裝置，其中該使用者介面係一彈出式視窗。
56. 如申請專利範圍第50項之裝置，其中如果該候選清單未包含該正確的目標字元字串的話，該裝置進一步包括：
用以由一使用者提供該目標字元字串的構件；
用以以手動的方式將該第一字元字串轉換成該目標字元字串的構件；以及
用以將該使用者所提供的該目標字元字串儲存於一資料庫中的構件。
57. 如申請專利範圍第56項之裝置，其中該目標字元字串係由該使用者透過一使用者介面進行輸入。
58. 如申請專利範圍第43項之裝置，其中該組規則係由一使用者透過該使用者介面加以定義。
59. 如申請專利範圍第33項之裝置，其中係透過對照表創造該第三字元字串。

60. 如申請專利範圍第33項之裝置，其中係透過演算法創造該第三字元字串。
61. 如申請專利範圍第40項之裝置，進一步包括分配一優先值給該等第三字元字串之每一個。
62. 如申請專利範圍第61項之裝置，其中該第三字元字串的優先值低於該第二字元字串的優先值。
63. 一種用以將一第一日文字元輸入字串轉換成一第二日文字元字串的裝置，該裝置包括：
- 用以接收一平假名輸入的構件；
 - 用以根據該平假名輸入而自動決定複數個可能的片假名候選項的構件；
 - 用以分析該複數個可能的片假名候選項的構件，以便將該平假名輸入轉換成片假名字元；
 - 依照該分析的結果選擇其中一個片假名候選項的構件；以及
 - 輸出包含該等片假名候選項的轉換內容的構件。
64. 如申請專利範圍第63項之裝置，其中該轉換內容進一步包括漢字字元。
65. 一種機器可讀取的媒體，其中儲存著可執行碼，用以讓機器執行將一原始字元字串轉換成一目標字元字串的方法，該方法包括：
- 接收具有該原始字元字串之第一字元字串；
 - 將該第一字元字串分割成複數個子字串；
 - 透過字典將該複數個子字串轉換成第二字元字串；

創造與該複數個子字串對應的第三字元字串；

分析該第二字元字串及該第三字元字串；

根據分析結果，由該第二及第三字元字串建立第四字元字串；

根據該第四字元字串創造一份候選清單；

從該候選清單中選出目標字元字串；以及

輸出該目標字元字串。

66. 如申請專利範圍第65項之機器可讀取的媒體，其中該第一字元字串包括日文平假名字元，並且其中在該接收之前，該字典係以一機器可讀取的形式儲存。

67. 如申請專利範圍第65項之機器可讀取的媒體，其中該等第二字元字串包括日文單詞。

68. 如申請專利範圍第65項之機器可讀取的媒體，其中該第三字元字串包括片假名字元。

69. 如申請專利範圍第65項之機器可讀取的媒體，其中該方法進一步包括：

如果該第二字元字串包含對應於該第一字元字串的一字元字串的話，便由該第二字元字串建立該第四字元字串；以及

如果該第二字元字串不包含對應於第一字元字串的字元字串的話，便由該第三字元字串建立該第四字元字串。

70. 如申請專利範圍第65項之機器可讀取的媒體，其中該方法進一步包括根據與該第二字元字串及該第三字元字串相關聯的優先資訊，對該第二字元字串及該第三字元字

串進行優先排序。

71. 如申請專利範圍第70項之機器可讀取的媒體，該優先資訊包括：

每個該第二字元字串的使用頻率；

字元分節長度；以及

該第二字元字串之間的連接關係資訊。

72. 如申請專利範圍第70項之機器可讀取的媒體，其中該方法進一步包括根據該優先資訊，計算該第二及第三字元字串中每一個的一優先值。

73. 如申請專利範圍第65項之機器可讀取的媒體，其中該方法進一步包括，從該字典中擷取出該第二字元字串的優先資訊。

74. 如申請專利範圍第65項之機器可讀取的媒體，其中該字典包括數個字典。

75. 如申請專利範圍第65項之機器可讀取的媒體，其中會根據一組規則分析該第二及第三字元字串。

76. 如申請專利範圍第75項之機器可讀取的媒體，其中該組規則包括下面的資訊：該些字元字串的使用頻率以及該些字元字串之間的連接關係。

77. 如申請專利範圍第76項之機器可讀取的媒體，其中該組規則進一步包括下面的資訊：字元分節長度以及語義/文法規則。

78. 如申請專利範圍第75項之機器可讀取的媒體，其中會根據該組規則，由該第二及第三字元字串建立該第四字元

字串。

79. 如申請專利範圍第65項之機器可讀取的媒體，其中該方法進一步包括計算每一個該第四字元字串之優先值。

80. 如申請專利範圍第79項之機器可讀取的媒體，其中會根據該第四字元字串的優先值創造該份候選清單。

81. 如申請專利範圍第80項之機器可讀取的媒體，其中會從該份候選清單中選出具最高優先值的候選單詞作為該目標字元字串。

82. 如申請專利範圍第65項之機器可讀取的媒體，其中該方法進一步包括：

檢查該目標字元字串的輸出結果，判斷該目標字元字串是否正確；

如果該輸出不正確的話，便擷取該替代字元字串候選清單；以及

從該候選清單中選擇一字元字串作為最終的輸出。

83. 如申請專利範圍第82項之機器可讀取的媒體，其中一使用者會檢查該目標字元字串的輸出。

84. 如申請專利範圍第82項之機器可讀取的媒體，其中該方法進一步包括將該候選清單顯示在一使用者介面中。

85. 如申請專利範圍第82項之機器可讀取的媒體，其中一使用者會從該候選清單中選擇一字元字串作為最終的輸出。

86. 如申請專利範圍第82項之機器可讀取的媒體，其中該方法進一步包括，根據從該候選清單中所選出的該最終輸

出，就該最終輸出字元字串的該優先資訊方面更新一資料庫。

87. 如申請專利範圍第84項之機器可讀取的媒體，其中該使用者介面係一彈出式視窗。

88. 如申請專利範圍第82項之機器可讀取的媒體，其中如果該候選清單未包含該正確的目標字元字串的話，該方法進一步包括：

由一使用者提供該目標字元字串；

以手動的方式將該第一字元字串轉換成該目標字元字串；以及

將該使用者所提供的該目標字元字串儲存於一資料庫中。

89. 如申請專利範圍第88項之機器可讀取的媒體，其中該目標字元字串係由該使用者透過一使用者介面進行輸入。

90. 如申請專利範圍第75項之機器可讀取的媒體，其中該組規則係由一使用者透過一使用者介面加以定義。

91. 如申請專利範圍第65項之機器可讀取的媒體，其中係透過對照表創造該第三字元字串。

92. 如申請專利範圍第65項之機器可讀取的媒體，其中係透過一演算法創造該第三字元字串。

93. 如申請專利範圍第72項之機器可讀取的媒體，其中該方法進一步包括，分配一優先值給該等第三字元字串之每一個。

94. 如申請專利範圍第93項之機器可讀取的媒體，其中該等

第三字元字串的優先值低於該等第二字元字串的優先值。

95. 一種用以將一第一日文字元輸入字串轉換成一第二日文字元字串的機器可讀取的媒體，該方法包括：

接收一平假名輸入；

根據該平假名輸入，自動決定複數個可能的片假名候選項；

分析該複數個可能的片假名候選項，以便將該平假名輸入轉換成片假名字元；

依照該分析的結果選擇其中一個片假名候選項；以及

輸出包含該個片假名候選項的轉換內容。

96. 如申請專利範圍第95項之機器可讀取的媒體，其中該轉換內容進一步包括漢字字元。

97. 一種用以將一原始字元字串轉換成一目標字元字串的裝置，其包括：

一輸入方法，用以接收具有該原始字元字串之該第一字元字串；

一一般字典，其係耦合成用以將該第一字元字串轉換成第二字元字串；

一虛擬字典，其係耦合成根據該第一字元字串以便產生第三字元字串；

一耦合至該輸入方法的型態分析引擎(MAE)，該MAE會對該第一字元字串進行型態分析，並且將根據該第二及第三字元字串，將該第一字元字串轉換成該目標字元

字串；

一耦合至該 MAE 的輸出單元。

98. 如申請專利範圍第 97 項之裝置，其中該輸入方法包括一應用程式介面 (API)，用以接收該第一字元字串。

99. 如申請專利範圍第 97 項之裝置，其中，該輸入方法係從一鍵盤接收該第一字元字串。

100. 如申請專利範圍第 97 項之裝置，其中該輸入方法係從一個人數位助理 (PDA) 接收該第一字元字串。

101. 如申請專利範圍第 97 項之裝置，其中該第一字元字串包括日文平假名字元。

102. 如申請專利範圍第 97 項之裝置，其中該目標字元字串包括日文字元，其中該日文字元包括：

平假名字元；

片假名字元；以及

漢字字元。

103. 如申請專利範圍第 97 項之裝置，其中該 MAE 會將該第一字元字串分割成複數個子字串，並且透過該一般字典將該複數個子字串轉換成該等第二字元字串。

104. 如申請專利範圍第 97 項之裝置，其中該 MAE 會將該第一字元字串分割成複數個子字串，並且透過該虛擬字典產生與該複數個子字串對應的該等第三字元字串。

105. 如申請專利範圍第 104 項之裝置，其中該虛擬字典提供該等第三字元字串的每一個部分語音資訊。

106. 如申請專利範圍第 104 項之裝置，其中該虛擬字典會分配

一優先值給每一個該第三字元字串。

107.如申請專利範圍第106項之裝置，其中，該第三字元字串的該優先值低於來自該一般字典的字元字串的該優先值。

108.如申請專利範圍第97項之裝置，其中，該一般字典包括數個字典。

109.如申請專利範圍第104項之裝置，其中，該MAE會從該一般字典的該輸出中選擇候選單詞，以及如果該一般字典並未包含該正確的轉換的話，該MAE便會從該虛擬字典的輸出中選擇該候選單詞。

110.如申請專利範圍第97項之裝置，進一步包括一耦合至該MAE的字典管理模組(DMM)，該字典管理係管理該一般字典及該虛擬字典。

111.如申請專利範圍第110項之裝置，其中該MAE會透過該DMM存取該一般字典及該虛擬字典。

112.如申請專利範圍第97項之裝置，進一步包括一使用者介面，其係耦合成用以對該MAE所產生的該轉換結果提供使用者互動。

113.如申請專利範圍第112項之裝置，其中如果該MAE所產生的該轉換結果不正確的話，一使用者便會從該MAE所產生的一候選清單中，透過該使用者介面選擇一最終的輸出。

114.如申請專利範圍第97項之裝置，其中該輸出單元會將該轉換結果傳輸給一顯示元件。

- 115.如申請專利範圍第97項之裝置，其中該輸出單元會透過一API將該轉換結果傳輸給一應用程式。
- 116.一種具有如申請專利範圍第97項之裝置的電腦系統。
- 117.如申請專利範圍第116項之電腦系統，其中該電腦系統係一網路電腦，而且其中該一般字典係儲存在網路儲存位置中。

拾壹、圖式

あ	い	う	え	お
か	き	く	け	こ
さ	し	す	せ	そ
た	ち	つ	て	と
な	に	ぬ	ね	の
は	ひ	ふ	へ	ほ
ま	み	む	め	も
や	ー	ゆ	ー	よ
ら	り	る	れ	ろ
わ	を	ん	ー	ー
か	き	く	け	こ
さ	し	す	せ	そ
た	ち	つ	て	と
は	ひ	ふ	へ	ほ
ば	び	ぶ	べ	ぼ

151

ア	イ	ウ	エ	オ
カ	キ	ク	ケ	コ
サ	シ	ス	セ	ソ
タ	チ	ツ	テ	ト
ナ	ニ	ヌ	ネ	ノ
ハ	ヒ	フ	ヘ	ホ
マ	ミ	ム	メ	モ
ヤ	ー	ユ	ー	ヨ
ラ	リ	ル	レ	ロ
ワ	ヲ	ン	ー	ー
ガ	ギ	グ	ゲ	ゴ
サ	シ	ス	セ	ソ
タ	チ	ツ	テ	ト
ハ	ヒ	フ	ヘ	ホ
バ	ビ	ブ	ベ	ボ

152

$\underbrace{\text{サンノセ}}_{154} \text{ } \overbrace{\text{映画を見る}}^{153}$

150

圖1A

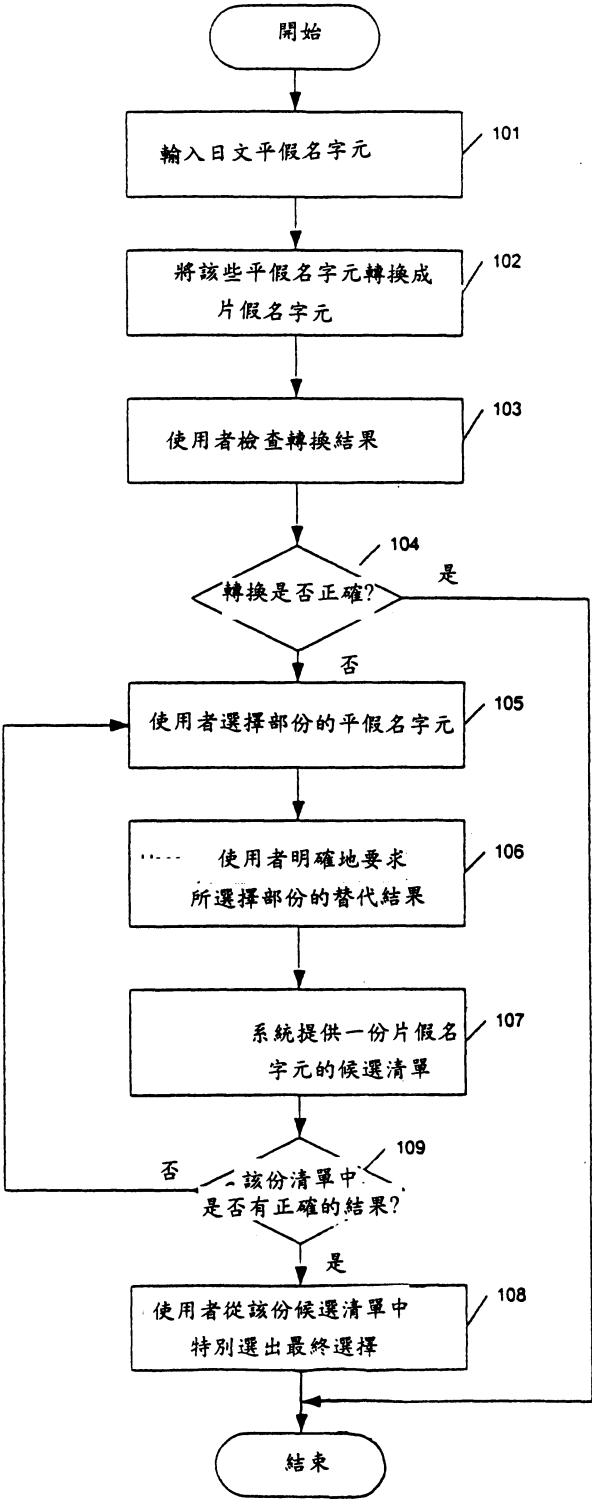


圖 1B

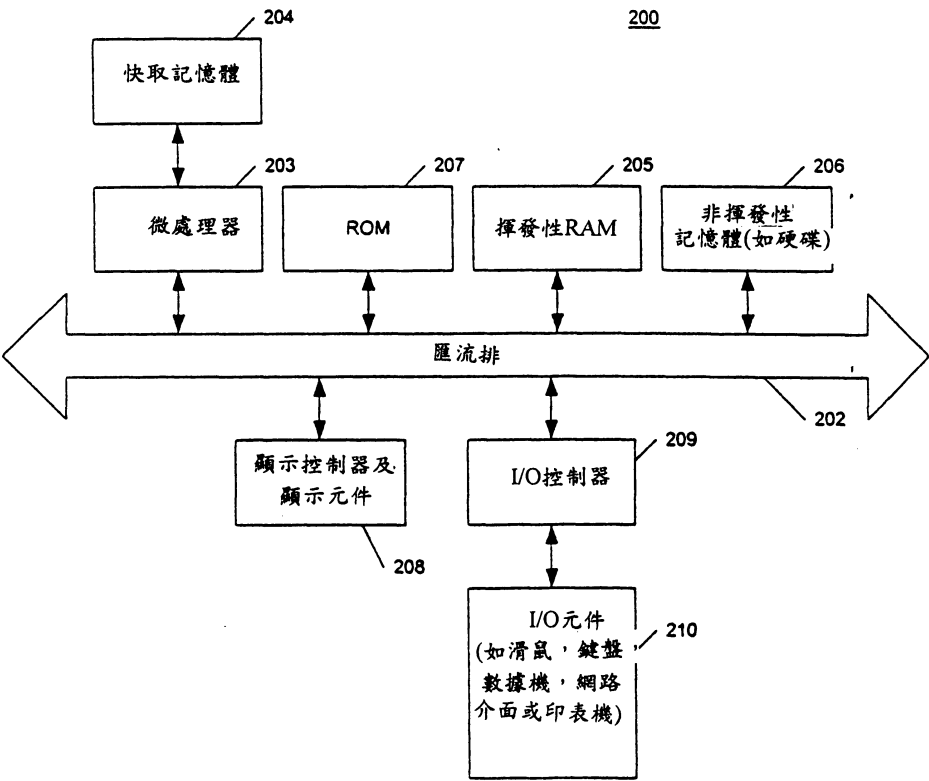


圖2

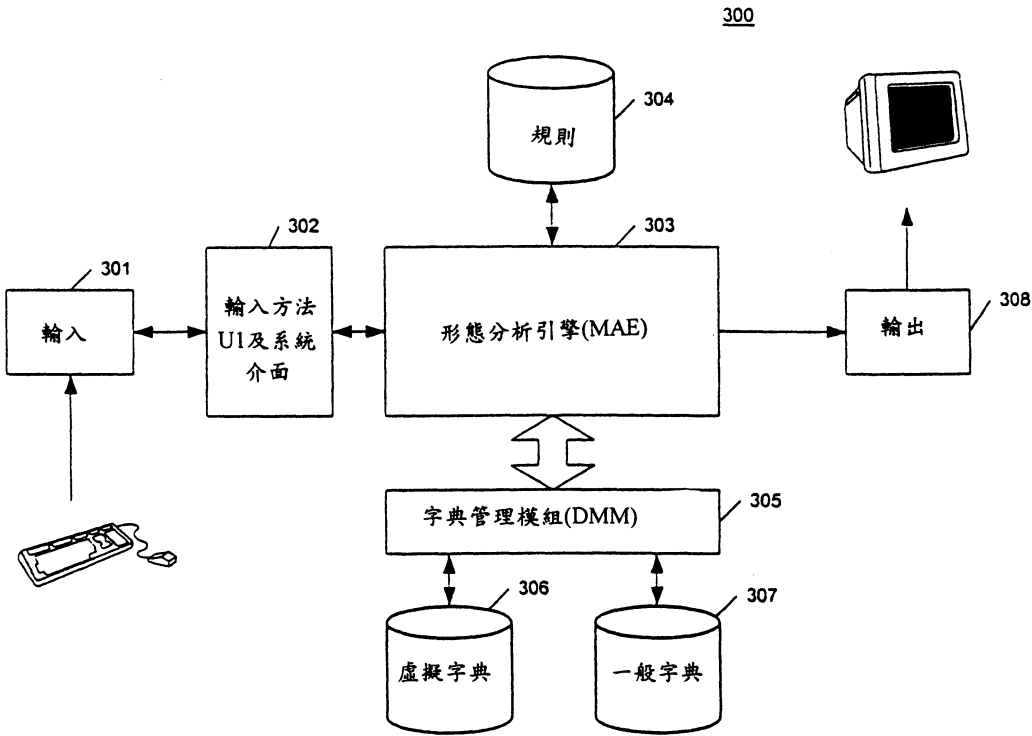


圖3

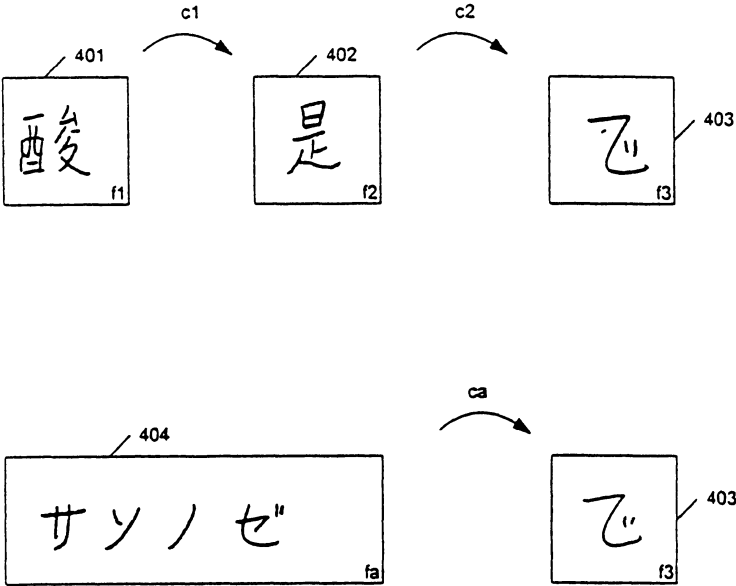


圖4

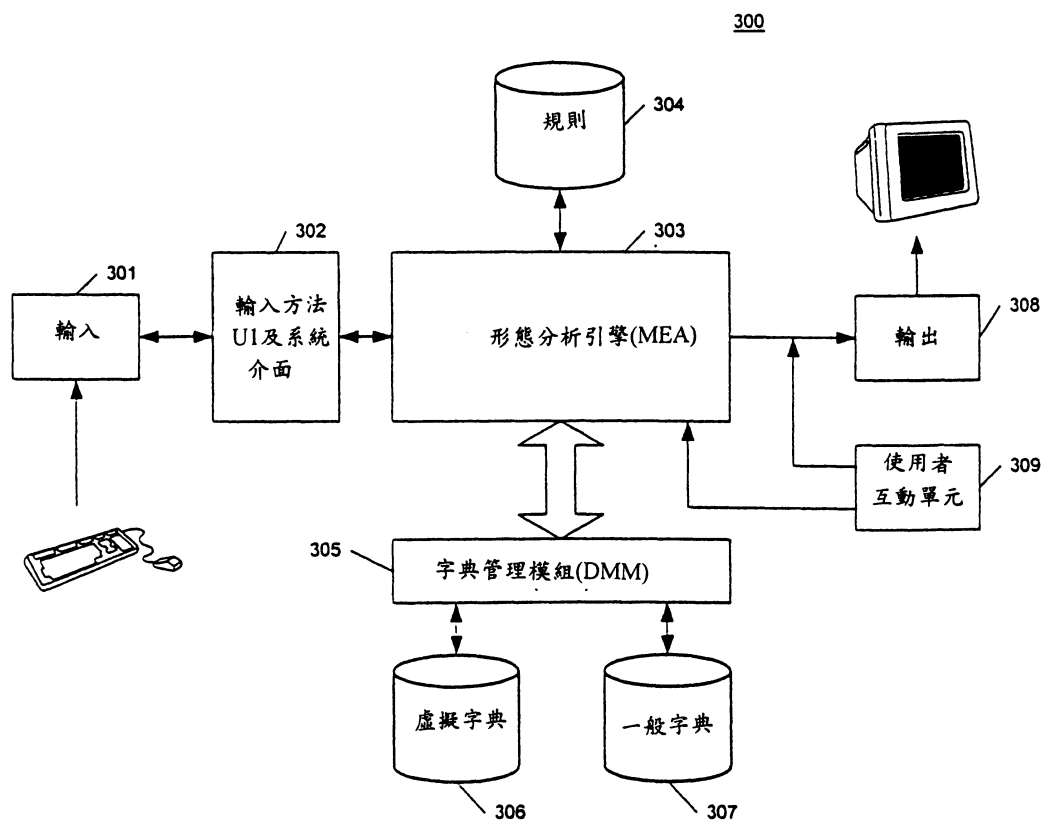


圖5

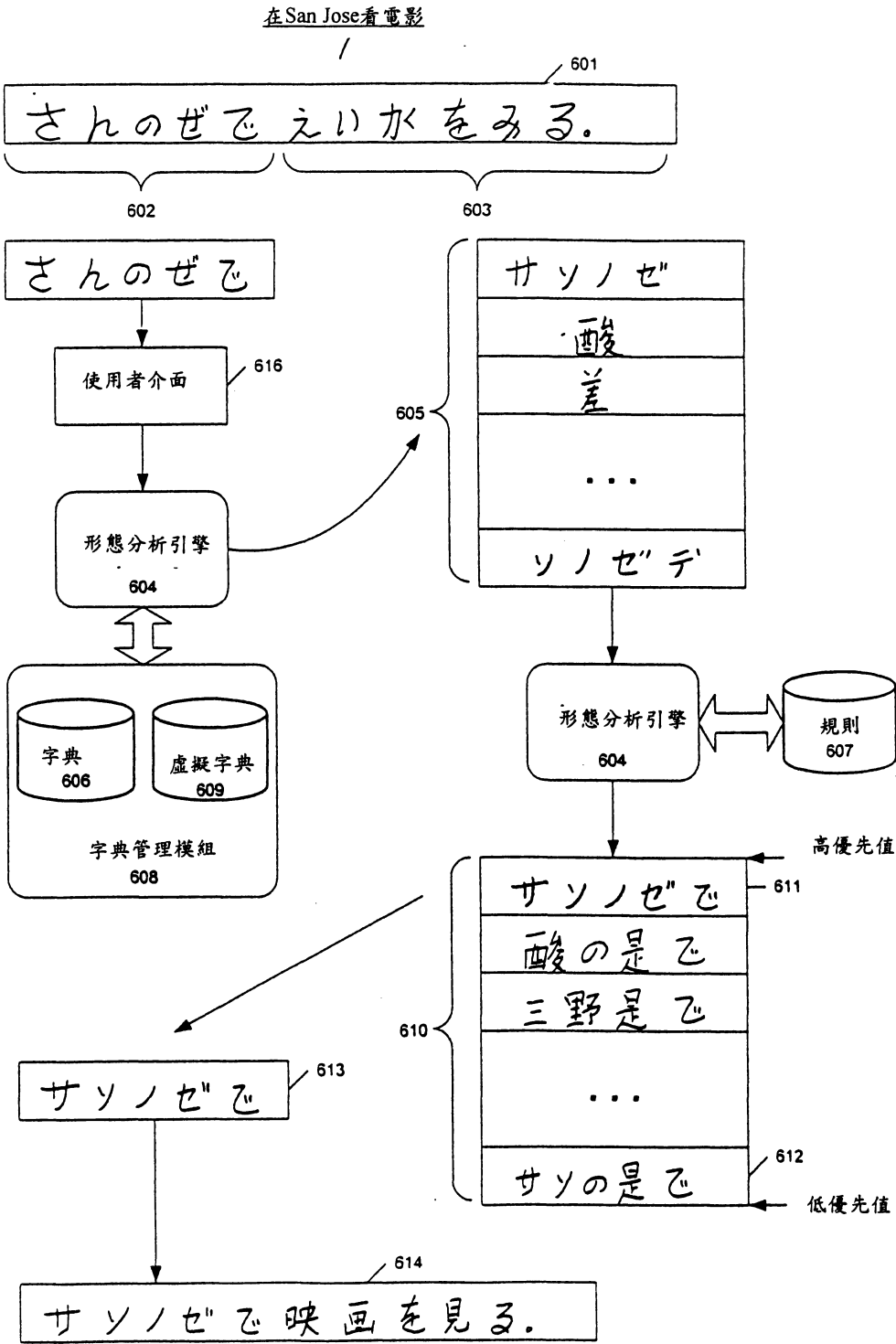


圖6A

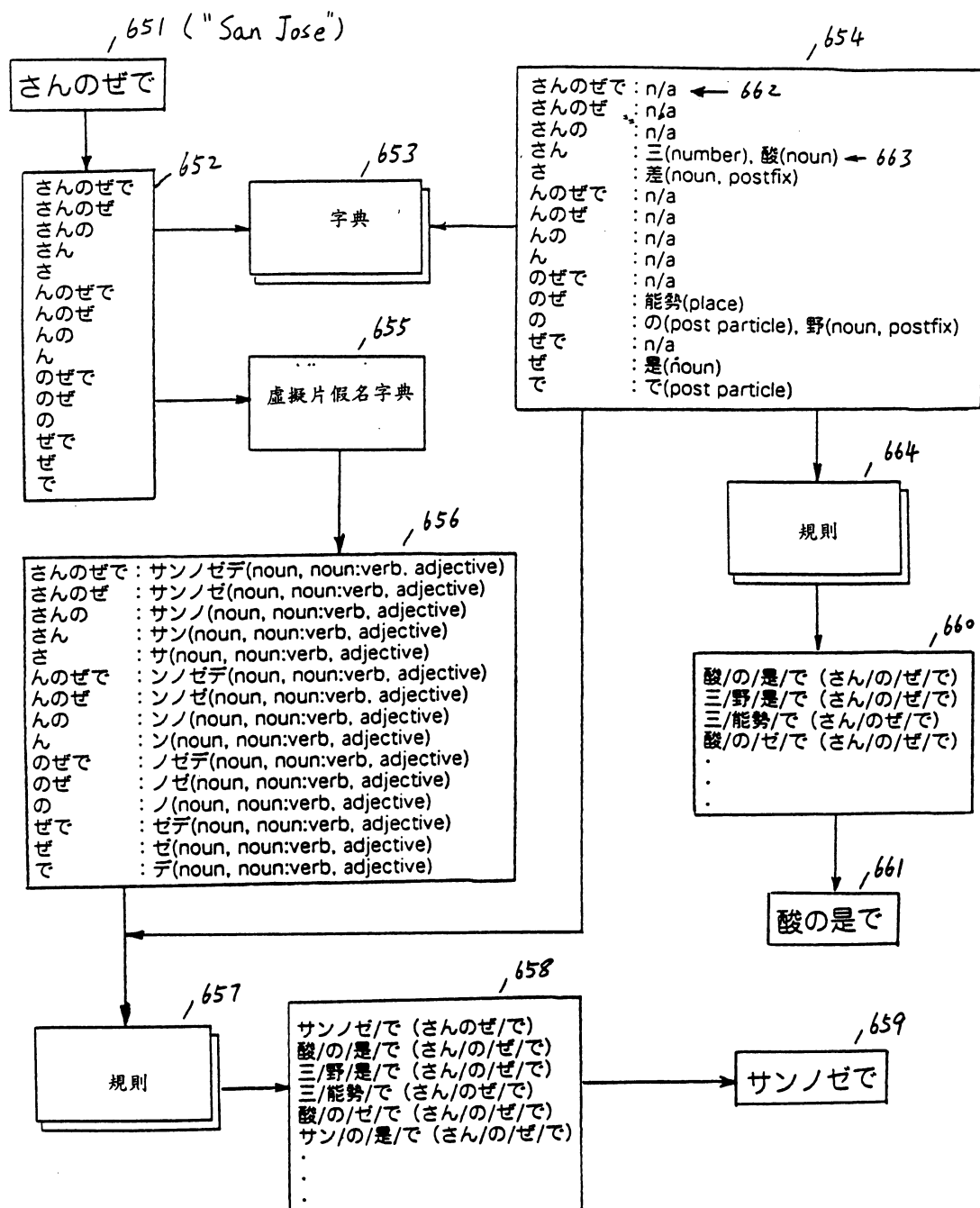


圖 6B

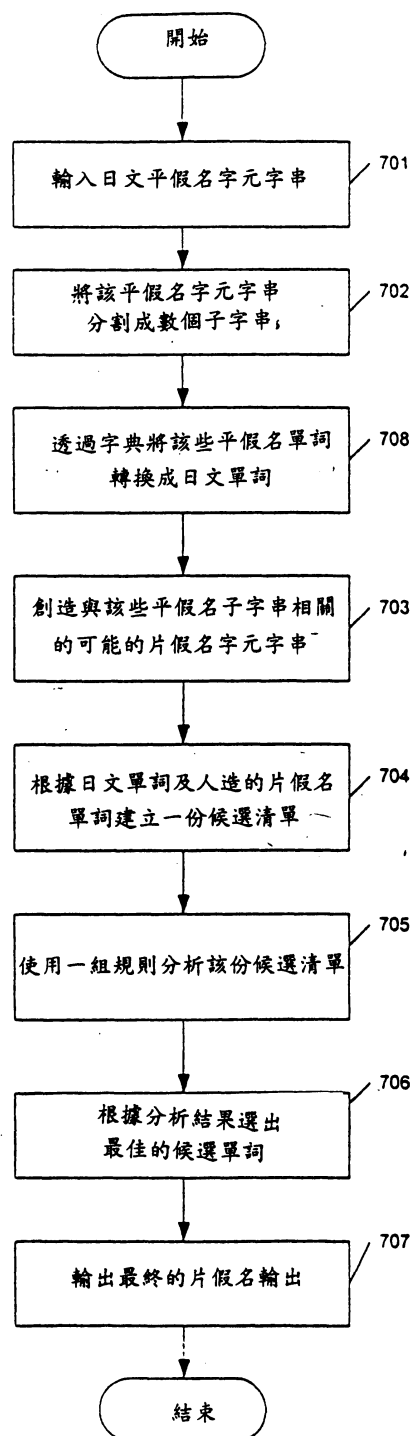


圖 7

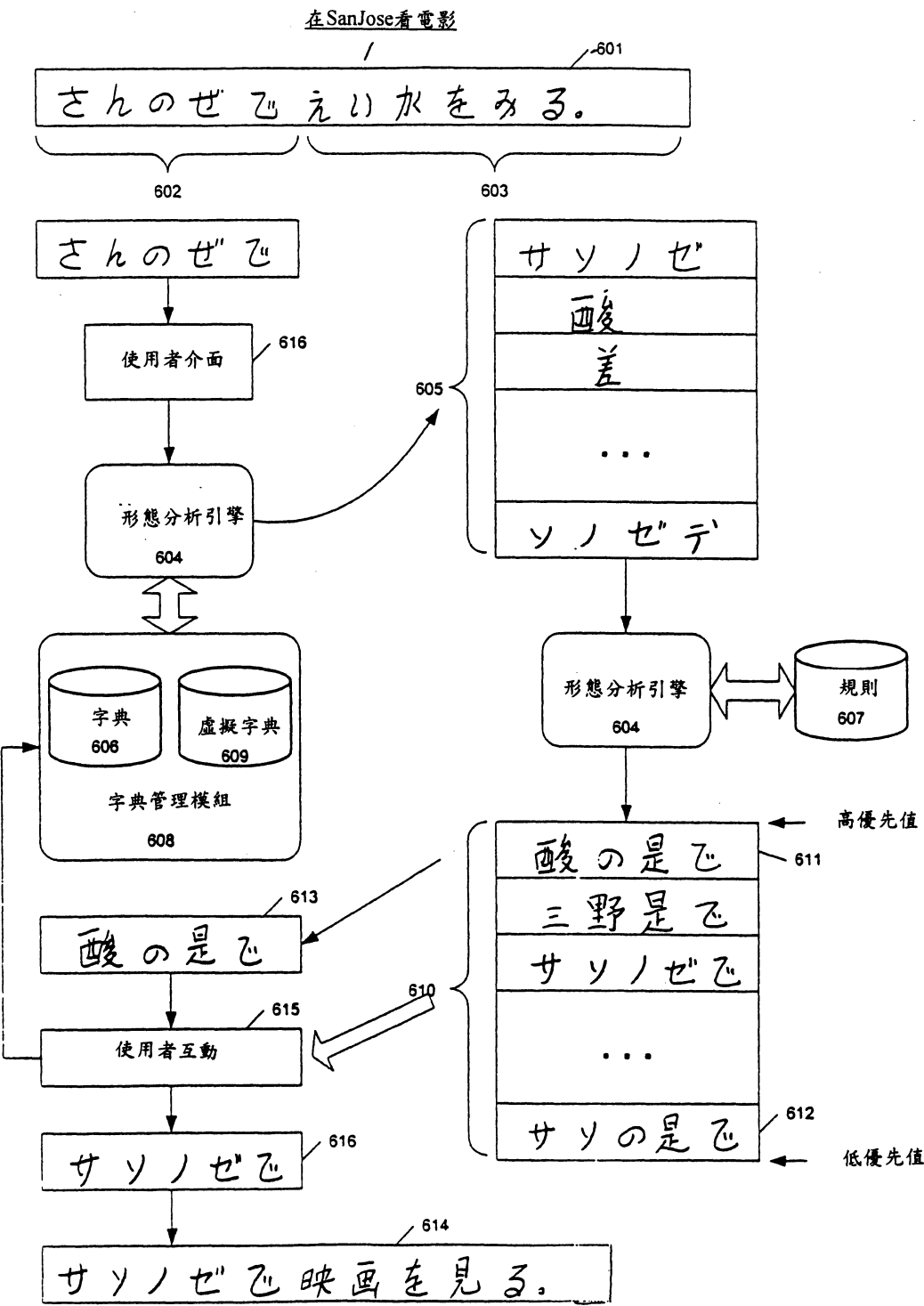


圖8

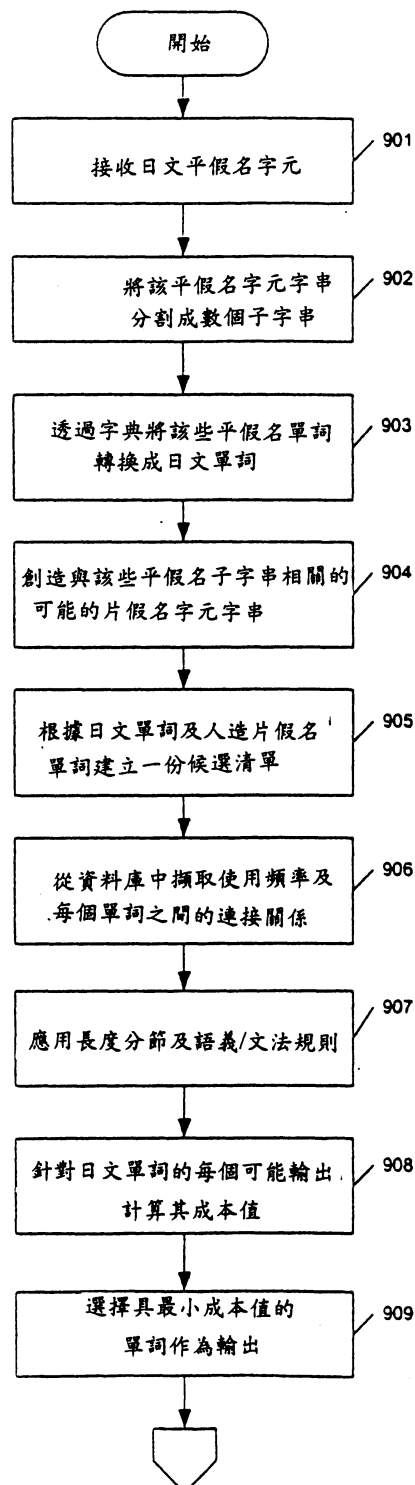


圖9A

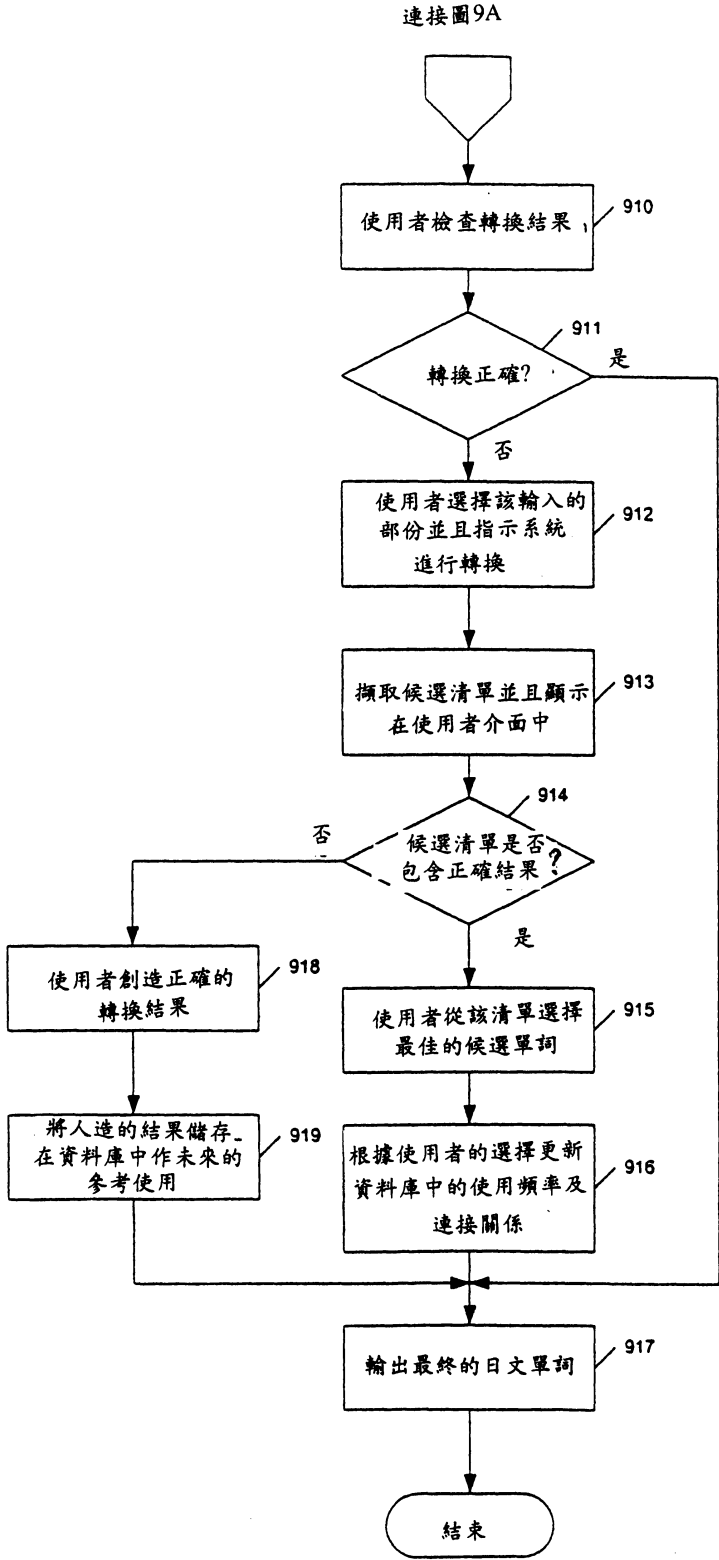


圖9B