



[12] 发明专利申请公开说明书

[21] 申请号 200410074943.5

[43] 公开日 2005年3月9日

[11] 公开号 CN 1591415A

[22] 申请日 2004.9.1
 [21] 申请号 200410074943.5
 [30] 优先权
 [32] 2003.9.1 [33] JP [31] 2003-308409
 [32] 2004.5.21 [33] JP [31] 2004-151965
 [71] 申请人 株式会社国际电气通信基础技术研究
 所
 地址 日本京都府
 [72] 发明人 渡边太郎 隅田英一郎

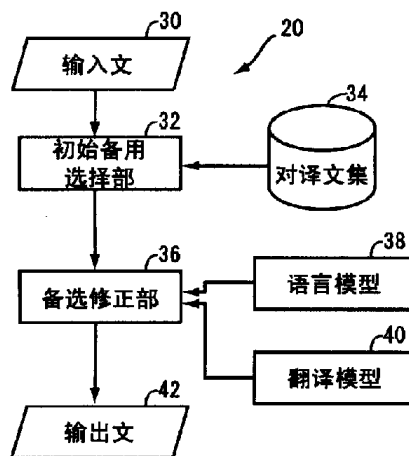
[74] 专利代理机构 中科专利商标代理有限责任公
 司
 代理人 汪惠民

权利要求书7页 说明书16页 附图6页

[54] 发明名称 机器翻译装置以及机器翻译计算机程序

[57] 摘要

本发明提供一种机器翻译方法，采用包含由第1种语言的句子和第2种语言的句子构成对译的对译文集，将第1种语言的输入文翻译为第2种语言的句子，其包括以下步骤：接收第1种语言的输入文，从对译文集之中，摘录出和输入文最类似的、和第1种语言的句子成对的第2种语言的句子的步骤；对于摘录出的第2种语言的句子，应用多个变形之中的任意的变形，计算由变形得到的句子的类似度的步骤；选择由变形得到的句子之中的、规定数目的、类似度高的句子的步骤；直到类似度没有改善为止，对于选择步骤中选择的句子的各个，重复摘录步骤、计算步骤以及选择步骤的步骤；当重复步骤结束时，将剩余的第2种语言的句子之中，具有最大类似度的句子作为对输入文的译文输出的步骤。



1、一种机器翻译方法，其采用包含多个由第 1 种语言的句子和第 2
5 种语言的句子构成对译的对译文集，将所述第 1 种语言输入文翻译为所述
第 2 种语言的句子，其特征在于，包括以下步骤：

摘录步骤，接收所述第 1 种语言输入文，从所述对译文集之中，摘录
出和所述输入文之间具有规定关系的、和第 1 种语言的句子成对的所述第
2 种语言的句子；

10 计算步骤，对于所述摘录出的所述第 2 种语言的句子，应用预先决定的
多个变形之中的任意变形，计算由变形得到的句子的类似度；

选择步骤，从由所述变形得到的句子中，选择具有满足规定条件的类
似度的句子；

重复步骤，直到预先决定的结束条件成立为止，对于所述选择步骤选
15 择的句子的每一个，重复所述摘录步骤、所述计算步骤以及所述选择步骤；
和

输出步骤，在所述重复步骤结束时剩余的所述第 2 种语言的句子之中，
具有满足预先决定的选择条件的类似度的句子作为对所述输入文的译文
并输出。

20 2、根据权利要求 1 所述的机器翻译方法，其特征在于，所述摘录步
骤包括读取步骤，其接收所述第 1 种语言的输入文，从所述对译文集之中，
读取当表示和所述输入文的类似度的规定分数满足预先决定的条件的、和
所述第 1 种语言的句子成对的所述第 2 种语言的句子。

25 3、根据权利要求 2 所述的机器翻译方法，其特征在于，所述读取步
骤包括：

分数计算步骤，其接收所述第 1 种语言的输入文，对于所述对译文集
之中所包含的所述第 1 种语言的句子的每一个，计算和所述输入文之间的
所述分数；

30 确定步骤，其确定在所述分数计算步骤中算出的分数最大的 1 个或者
多个的所述第 1 种语言的句子；和

句子读出步骤，其将所述确定步骤中决定的所述 1 个或者多个第 1 种语言的句子和分别与其成对的 1 个或者多个所述第 2 种语言的句子从所述对译文集之中读出。

4、根据权利要求 3 所述的机器翻译方法，其特征在于，所述分数计算步骤包括：

类似尺度计算步骤，其在所述输入文和所述对译文集之中所包含的所述第 1 种语言的各个之间，将所述对译文集之中所包含的所述第 1 种语言的句子作为文本，对所述输入文采用定义的文本频率计算规定的类似尺度；

10 编辑距离计算步骤，其计算所述输入文、与包含在所述对译文集之中的所述第 1 种语言的各个之间的编辑距离；和

分数算出步骤，其基于在所述类似尺度计算步骤中算出的类似尺度以及在所述编辑距离计算步骤中算出的编辑距离，计算所述分数。

5、根据权利要求 4 所述的机器翻译方法，其特征在于，所述类似尺度计算步骤包括 tf/idf 标准 $P_{tf/idf}$ 计算步骤，在所述对译文集之中所包含的所述第 1 种语言的各句子和所述输入文之间，按照下面的公式计算 tf/idf 标准 $P_{tf/idf}$ ，

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

20 式中， J_0 表示输入文， $J_{0,i}$ 表示输入文 J_0 的第 i 个单词， $df(J_{0,i})$ 表示对于单词 $J_{0,i}$ 的文本频率， J_k 表示第 k 个所述第 1 种语言的句子， $1 \leq k \leq N$ ， N 表示对译文集之中的全部对译文数目。

6、根据权利要求 5 所述的机器翻译方法，其特征在于，所述编辑距离计算步骤包括：进行输入文 J_0 和所述第 1 种语言的句子 J_k 之间的 DP 匹配，即动态编程匹配，计算编辑距离 $dis(J_k, J_0)$ 的步骤，编辑距离 dis 25 (J_k, J_0) 由以下公式决定，

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

式中， k 是 $1 \leq k \leq N$ 的整数， $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 以及 $S(J_k, J_0)$ 分别为将句子 J_0 变换为句子 J_k 时，必要的词语的插入、删除以及置换的数量。

7、根据权利要求 6 所述的机器翻译方法，其特征在于，所述分数算

出步骤包括：

分数求出步骤，其对所述第 1 种语言的句子 J_k ，基于在所述类似尺度计算步骤中算出的 tf/idf 标准 $P_{tf/idf}$ 以及在所述编辑距离计算步骤算出的编辑距离 $dis(J_k, J_0)$ ，求出由下面公式所定义的分数的

$$5 \quad score = \begin{cases} (1.0 - \alpha)(1.0 - \frac{dis(J_k, J_0)}{|J_0|}) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ 时}) \\ 1.0 & (\text{其它情况}) \end{cases}$$

式中， α 为调整参数；

对译文选择步骤，其从在所述分数求出步骤中求得的分数较大的对译文中，作为所述初始备用，依次选择预先决定个数的对译文。

8、根据权利要求 7 所述的机器翻译方法，其特征在于，进一步包括：
10 判定步骤，其判断在所述读取步骤中读出的所述第 1 种语言的句子中是否存在所述分数为 1 的句子；和

译文输出步骤，其对在所述判定步骤中判断为存在分数为 1 的所述第 1 种语言的句子一事进行应答，将该分数为 1 的所述第 1 种语言的句子作为对于所述输入文的译文输出。

9、根据权利要求 7 所述的机器翻译方法，其特征在于，进一步包括：
15 对在所述判定步骤中判断存在分数为 1 的所述第 2 种语言的句子一事进行应答，对所述计算步骤，所述选择步骤，所述重复步骤以及所述输出步骤的执行进行抑制的步骤。

10、根据权利要求 1 所述的机器翻译方法，其特征在于，所述选择步
20 骤包括：

在由所述变形得到的句子中，选择规定个数的类似度最高的句子的步骤。

11、根据权利要求 1 所述的机器翻译方法，其特征在于，所述重复步
骤包括：

25 直到看不出所述选择步骤所选择的句子的类似度改善为止，对于所述选择步骤中选择的句子的各个，重复所述摘录步骤、所述计算步骤以及所述选择步骤的步骤。

12、根据权利要求 1 所述的机器翻译方法，其特征在于，所述输出步

骤包括：

在所述重复步骤结束时剩余的所述第2种语言的句子之中，具有最大的类似度的句子作为对于所述输入文的译文输出的步骤。

13、根据权利要求1所述的机器翻译方法，其特征在于，

5 所述机器翻译方法与所述第2种语言的语言模型、以及由所述第2种语言向所述第1种语言的翻译模型连接并使用，

所述计算步骤包括：对于摘录出的所述第2种语言的句子，应用所述预先决定的多个变形之中的任意变形，采用所述语言模型以及所述翻译模型计算由变形得到的句子的类似度的步骤。

10 14、一种记录介质，记录了让计算机动作的机器翻译计算机程序，在计算机执行该程序时将实施以下机器翻译方法，所述机器翻译方法采用包含多个由第1种语言的句子和第2种语言的句子构成对译的对译文集，将所述第1种语言输入文翻译为所述第2种语言，其特征在于，包括以下步骤：

15 摘录步骤，接收所述第1种语言输入文，从所述对译文集之中，摘录出和所述输入文之间具有规定关系的、和第1种语言的句子成对的所述第2种语言的句子；

计算步骤，对于所述摘录出的所述第2种语言的句子，应用预先决定的多个变形之中的任意变形，计算由变形得到的句子的类似度；

20 选择步骤，从由所述变形得到的句子中，选择具有满足规定条件的类似度的句子；

重复步骤，直到预先决定的结束条件成立为止，对于所述选择步骤选择的句子的每一个，重复所述摘录步骤、所述计算步骤以及所述选择步骤；
和

25 输出步骤，在所述重复步骤结束时剩余的所述第2种语言的句子之中，具有满足预先决定的选择条件的类似度的句子作为对所述输入文的译文并输出。

15、根据权利要求14所述的记录介质，其特征在于，所述摘录步骤包括读取步骤，其接收所述第1种语言的输入文，从所述对译文集之中，
30 读取当表示和所述输入文的类似度的规定分数满足预先决定的条件的、和

所述第 1 种语言的句子成对的所述第 2 种语言的句子。

16、根据权利要求 15 所述的记录介质，其特征在于，所述读取步骤包括：

5 分数计算步骤，其接收所述第 1 种语言的输入文，对于所述对译文集之中所包含的所述第 1 种语言的句子的每一个，计算和所述输入文之间的所述分数；

确定步骤，其确定在所述分数计算步骤中算出的分数最大的 1 个或者多个的所述第 1 种语言的句子；和

10 句子读出步骤，其将所述确定步骤中决定的所述 1 个或者多个第 1 种语言的句子和分别与其成对的 1 个或者多个所述第 2 种语言的句子从所述对译文集之中读出。

17、根据权利要求 16 所述的记录介质，其特征在于，所述分数计算步骤包括：

15 类似尺度计算步骤，其在所述输入文和所述对译文集之中所包含的所述第 1 种语言的各个之间，将所述对译文集之中所包含的所述第 1 种语言的句子作为文本，对所述输入文采用定义的文本频率计算规定的类似尺度；

编辑距离计算步骤，其计算所述输入文、与包含在所述对译文集之中的所述第 1 种语言的各个之间的编辑距离；和

20 分数算出步骤，其基于在所述类似尺度计算步骤中算出的类似尺度以及在所述编辑距离计算步骤中算出的编辑距离，计算所述分数。

18、根据权利要求 17 所述的记录介质，其特征在于，所述类似尺度计算步骤包括 tf/idf 标准 $P_{tf/idf}$ 计算步骤，在所述对译文集之中所包含的所述第 1 种语言的各句子和所述输入文之间，按照下面的公式计算 tf/idf 标
25 准 $P_{tf/idf}$ ，

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

式中， J_0 表示输入文， $J_{0,i}$ 表示输入文 J_0 的第 i 个单词， $df(J_{0,i})$ 表示对于单词 $J_{0,i}$ 的文本频率， J_k 表示第 k 个所述第 1 种语言的句子， $1 \leq k \leq N$ ， N 表示对译文集之中的全部对译文数目。

19、根据权利要求 14 所述的记录介质，其特征在于，所述编辑距离计算步骤包括：进行输入文 J_0 和所述第 1 种语言的句子 J_k 之间的 DP 匹配，即动态编程匹配，计算编辑距离 $dis(J_k, J_0)$ 的步骤，编辑距离 $dis(J_k, J_0)$ 由以下公式决定，

$$5 \quad dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

式中， k 是 $1 \leq k \leq N$ 的整数， $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 以及 $S(J_k, J_0)$ 分别为将句子 J_0 变换为句子 J_k 时，必要的词语的插入、删除以及置换的数量。

20、根据权利要求 19 所述的记录介质，其特征在于，所述分数算出步骤包括：

10 分数求出步骤，其对所述第 1 种语言的句子 J_k ，基于在所述类似尺度计算步骤中算出的 tf/idf 标准 $P_{tf/idf}$ 以及在所述编辑距离计算步骤算出的编辑距离 $dis(J_k, J_0)$ ，求出由下面公式所定义的分数的分数，

$$score = \begin{cases} (1.0 - \alpha) \left(1.0 - \frac{dis(J_k, J_0)}{|J_0|}\right) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ 时}) \\ 1.0 & (\text{其它情况}) \end{cases}$$

式中， α 为调整参数；

15 对译文选择步骤，其从在所述分数求出步骤中求得的分数的较大的对译文中，作为所述初始备用，依次选择预先决定个数的对译文。

21、根据权利要求 20 所述的记录介质，其特征在于，所述机器翻译方法进一步包括：

20 判定步骤，其判断在所述读取步骤中读出的所述第 1 种语言的句子中是否存在所述分数为 1 的句子；和

译文输出步骤，其对在所述判定步骤中判断为存在分数为 1 的所述第 1 种语言的句子一事进行应答，将该分数为 1 的所述第 1 种语言的句子作为对于所述输入文的译文输出。

25 22、根据权利要求 21 所述的记录介质，其特征在于，所述机器翻译方法进一步包括：

对在所述判定步骤中判断存在分数为 1 的所述第 2 种语言的句子一事进行应答，对所述计算步骤，所述选择步骤，所述重复步骤以及所述输出步骤的执行进行抑制的步骤。

23、根据权利要求 14 所述的记录介质，其特征在于，所述选择步骤包括：

在由所述变形得到的句子中，选择规定个数的类似度最高的句子的步骤。

5 24、根据权利要求 14 所述的记录介质，其特征在于，所述重复步骤包括：

直到看不出所述选择步骤所选择的句子的类似度改善为止，对于所述选择步骤中选择的句子的各个，重复所述摘录步骤、所述计算步骤以及所述选择步骤的步骤。

10 25、根据权利要求 14 所述的记录介质，其特征在于，所述输出步骤包括：

在所述重复步骤结束时剩余的所述第 2 种语言的句子之中，具有最大的类似度的句子作为对于所述输入文的译文输出的步骤。

26、根据权利要求 14 所述的记录介质，其特征在于，

15 所述机器翻译方法与所述第 2 种语言的语言模型、以及由所述第 2 种语言向所述第 1 种语言的翻译模型连接并使用，

所述计算步骤包括：对于摘录出的所述第 2 种语言的句子，应用所述预先决定的多个变形之中的任意变形，采用所述语言模型以及所述翻译模型计算由变形得到的句子的类似度的步骤。

20

机器翻译装置以及机器翻译计算机程序

5

技术领域

本发明涉及一种机器翻译装置，特别是在发挥范例翻译的优点的同时能进行高精度的翻译的统计机器翻译装置。

10 背景技术

在统计性的翻译中，将某一种语言的句子（J）翻译为其它语言的句子（E）的问题，作为最大化以下附带条件的概率的问题被定型化。

$$\hat{E} = \arg \max_E P(E|J)$$

对于此公式，通过适用贝叶斯定理得到下面的公式。

$$15 \quad \hat{E} = \arg \max_E P(E)P(J|E) / P(J)$$

其中，P(J)与 \hat{E} 的计算无关。因此，能得到下面的公式。

$$\hat{E} = \arg \max_E P(E)P(J|E)$$

右边的第1项P(E)被称为语言模型，表示句子E的类似度。第2项的P(J|E)被称为翻译模型，表示从句子E生成句子J的概率。

20 在这样的思考方式下，由单词排列（单词的对应）这样的概念提出了所谓使第1种语言的句子（称为信道目标文）映射为第2种语言的句子（称为信道源文）的翻译模型。此翻译模型在法语和英语、德语和英语等相互间类似的语言之间获得了成功。

25 但是，此翻译模型例如应用在日语和英语等相互间差异很大的语言时没有取得成果。这样的结果是因为在构造相互不同的语言间进行映射时，频繁地进行单词的插入和删除，各单词的派生很多以及词语的对应很复杂等的原因，带来了搜索空间庞大这样的问题而产生的。因为搜索变复杂，所以如果采用根据束搜索（beam search）的解码算法，那么在多数的情况

下只能得到局部的解。

基于单词排列 (alignment) 的统计的机器翻译采用单词排列 A 这样的考虑方法表示 2 种语言的对应。此时, 允许 1 个单词和多个单词相对应。所谓单词排列 A 是指将信道目标文的各个单词和信道源文的哪个单词相对应, 采用信道源文的单词指标进行表示的排列。在此排列中, 和信道源文的单词的对应采用赋予信道源文的单词的指标表示, 这些指标根据信道目标文的单词的顺序排列。

图 7 表示英语 (E) 以及日语 (J) 句子之间的单词排列的例 A。参照图 7, 将第 2 种语言 (此处为英语 E) 的句子 110 的 1~7 的各个单词, 和第 1 种语言 (此处为日语 J) 的句子 114 的各个单词 1~6 对应。对应由连接信道源文 110 的各个单词和信道目标文 114 的单词的连线群 112 表示。例如, 由信道源文 110 的单词[show₁]生成信道目标文 114 的两个单词[mise₅] (让看) 以及[tekudasai₆] (请)。另外信道目标文 114 的两个单词[no₂] (的) 以及[o₄] (を) 和信道源文 110 的哪个单词都没有对应, 假设信道源文 110 的开始部分为[null₀] (空), 作为与此对应的单词。这样的话此时的排列 A 为[7, 0, 4, 0, 1, 1]。

作为单词排列如果假设为这样的映射, 翻译模型 P (J|E) 能进一步写成以下这样。

$$P(J|E) = \sum_A P(J, A|E)$$

右边的项 P (J|E) 被进一步分解为、4 个要素, 由此 4 个要素构成将信道源文 E 变换为具有排列 A 的信道目标文 J 的以往技术的顺序。此 4 个要素如下。

(1) 根据派生模式, 对于信道源文的各单词决定能生成几个翻译词。有由一个单词可以生成两个翻译词的情况, 也有一个翻译词都不能生成的情况。

(2) 根据 NULL 生成模式, 在信道源文的适当位置插入 NULL。

(3) 通过查找词汇模型, 进行对生成的各单词的翻译。

(4) 通过参照变形模型, 对翻译的各单词进行排列变换。为了保存有关句子的限制条件, 此时的位置由刚才的单词排列决定。

这样，基于单词排列这样的思考方式得到翻译模型。

另一方面，提出了最初生成按照信道目标文的顺序排列的、由最初信道目标文的各单词翻译为信道源文的语言的句子，对于此文字应用各种运算符生成多个译文的方法。（Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu, Kenji Yamada:“机器翻译的快速解码以及优化解码”（2001）ACL2001 会议录，图卢兹，法国）。在此提出的方法中，在这样生成的译文之中，求出类似度最高的句子作为翻译文选择出来。

基于单词排列的统计翻译模型是对于作为例如法语和英语等相互类似的语言的两种语言设计的。另一方面，日语和英语相互之间具有极其不同的构造。因此，日语和英语相互翻译的情况下，图 7 所示那样的单词排列就变得非常复杂。此复杂度反映了语言的构造的差异。例如对于英语采用 SVO 的句型，而通常日语采用 SOV 这样的句型。还有，从图 7 所示的例子中也可以看出，非常频繁地产生插入和删除。例如对于图 7 所示的[the₃]以及[the₆]，在日语中不存在对应的形态要素。也就是说，对于这样的词当从日语翻译为英语时，必须进行插入。同样，对于日语的 no₂ 以及 o₄ 也必须进行删除。

由于这样的排列复杂以及词语的插入和删除频繁地进行，如果对每个单词进行束搜索，会产生计算量增大这样的问题。为了能在一定的时间内得出结果，就必须进行某些形式的修改。但是，通常的搜索算法中，如果这样在限定的空间中进行搜索的话，不可避免的会产生搜索误差。我们承认翻译的质量和由翻译模型指定的类似度之间存在某种程度的相关性，但由束搜索得到更高的质量是困难的。

另外在 Germann 等的方法中，存在着在搜索中得到局部的最适当的解很多这样的问题，不能稳定得到高精度的解。

25

发明内容

本发明的目的在于提供一种采用统计机器翻译的机器翻译方法及其装置，与语言的组合无关，而可以得到高品质的翻译。

本发明的另一目的在于提供一种采用统计机器翻译的机器翻译方法及其装置，其与语言的组合无关，而可以在一定程度的时间内得到高品质

30

的翻译。

本发明的又一目的在于提供一种采用统计机器翻译的机器翻译方法及其装置，其与语言的组合无关，而可以稳定得到高品质的翻译。

有关本发明第1方面的机器翻译方法，其采用包含多个由第1种语言的句子和第2种语言的句子构成对译的对译文集，将第1种语言输入文翻
5 译为第2种语言的句子，包括以下步骤：摘录步骤，接收第1种语言输入文，从对译文集之中，摘录出和输入文之间具有规定关系的、和第1种语言的句子成对的第2种语言的句子；计算步骤，对于摘录出的第2种语言的句子，应用预先决定的多个变形之中的任意变形，计算由变形得到的句
10 子的类似度；选择步骤，从由变形得到的句子中，选择具有满足规定条件的类似度的句子；重复步骤，直到预先决定的结束条件成立为止，对于选择步骤选择的句子的每一个，重复摘录步骤、计算步骤以及选择步骤；和输出步骤，在重复步骤结束时剩余的第2种语言的句子之中，具有满足预先决定的选择条件的类似度的句子作为对输入文的译文并输出。

15 对于输入文，从对译文集之中摘录出规定的关系成立的、和第1种语言成对的第2种语言的句子。对此第2种语言的句子进行种种的变形，重复从得到的句子中选择具有满足规定的条件的类似度的句子，将最终满足选择条件的句子作为对于输入文的译文输出。因为在对译文集之中的对译文是两种语言之间相互较好的对译文，所以摘录出的第2种语言的句子和
20 输入文的理想的译文相类似的可能性很高。重复这样摘录出的第2种语言的句子的种种变形，从得到的句子中基于类似度选择的译文，成为输入文的理想的译文的可能性很高。因为考虑最初摘录出的句子与理想的译文接近，所以在重复的过程中限于局部的最适当解的危险性很低。

优选，摘录步骤包括读取步骤，其接收第1种语言的输入文，从对译文集之中，读取当表示和输入文的类似度的规定分数满足预先决定的条件的、和第1种语言的句子成对的第2种语言的句子。
25

优选，读取步骤包括：分数计算步骤，其接收第1种语言的输入文，对于对译文集之中所包含的第1种语言的句子的每一个，计算和输入文之间的分数；确定步骤，其确定在分数计算步骤中算出的分数最大的1个或者多个的第1种语言的句子；和句子读出步骤，其将确定步骤中决定的1
30

个或者多个第 1 种语言的句子和分别与其成对的 1 个或者多个第 2 种语言的句子从对译文集之中读出。

每当摘录出第 2 种语言的句子时，确定表示和输入文的类似度的分数最大的 1 个或者多个第 1 种语言的句子，将这些句子和与其分别对应的第 2 种语言的句子从对译文集之中读出。以读出的第 2 种语言的句子为种子，进行变形以及类似度计算的循环，将得到的句子中满足规定的条件的句子作为输入文的译文。得到的第 2 种语言的句子对于输入文来讲，和理想的译文类似的可能性高，最终得到的译文相对于输入文来讲成为理想的译文的可能性增高。

分数计算步骤也可以包括：类似尺度计算步骤，其在输入文和对译文集之中所包含的第 1 种语言的各个之间，将对译文集之中所包含的第 1 种语言的句子作为文本，对输入文采用定义的文本频率计算规定的类似尺度；编辑距离计算步骤，其计算输入文、与包含在对译文集之中的第 1 种语言的各个之间的编辑距离；和分数算出步骤，其基于在类似尺度计算步骤中算出的类似尺度以及在编辑距离计算步骤中算出的编辑距离，计算分数。

优选，类似尺度计算步骤包括 tf/idf 标准 $P_{tf/idf}$ 计算步骤，在对译文集之中所包含的第 1 种语言的各句子和输入文之间，按照下面的公式计算 tf/idf 标准 $P_{tf/idf}$ ，

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

式中， J_0 表示输入文， $J_{0,i}$ 表示输入文 J_0 的第 i 个单词， $df(J_{0,i})$ 表示对于单词 $J_{0,i}$ 的文本频率， J_k 表示第 k 个第 1 种语言的句子 ($1 \leq k \leq N$)， N 表示对译文集之中的全部对译文数目。

进一步，编辑距离计算步骤包括：进行输入文 J_0 和第 1 种语言的句子 J_k 之间的 DP 匹配，即动态编程匹配，计算编辑距离 $dis(J_k, J_0)$ 的步骤，编辑距离 $dis(J_k, J_0)$ 由以下公式决定，

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

式中， k 是 $1 \leq k \leq N$ 的整数， $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 以及 $S(J_k, J_0)$ 分别为将句子 J_0 变换为句子 J_k 时，必要的词语的插入、删除以及置换的数量。

分数算出步骤包括：分数求出步骤，其对第 1 种语言的句子 J_k ，基于在类似尺度计算步骤中算出的 tf/idf 标准 $P_{tf/idf}$ 以及在编辑距离计算步骤算出的编辑距离 $dis(J_k, J_0)$ ，求出由下面公式所定义的分数的，

$$score = \begin{cases} (1.0 - \alpha)(1.0 - \frac{dis(J_k, J_0)}{|J_0|}) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ 时}) \\ 1.0 & (\text{其它情况}) \end{cases}$$

- 5 式中， α 为调整参数；对译文选择步骤，其在分数求出步骤中求得的分数的较大的对译文中，作为初始备用，依次选择预先决定个数的对译文。

机器翻译方法也可以进一步包括：判定步骤，其判断在读取步骤中读出的第 1 种语言的句子中是否存在分数为 1 的句子；和译文输出步骤，其对在判定步骤中判断为存在分数为 1 的第 1 种语言的句子一事进行应答，
10 将该分数为 1 的第 1 种语言的句子作为对于输入文的译文输出。

分数为 1 是指在对译文集之中存在和输入文一致的第 1 种语言的句子。即通过将此和第 1 种语言的句子成对的第 2 种语言的句子作为对于输入文的译文进行选择，能够得到很好的译文。

重复步骤也可以包括：直到看不出选择步骤所选择的句子的类似度改善为止，对于选择步骤中选择的句子的各个，重复摘录步骤、计算步骤以及选择步骤的步骤。
15

优选，输出步骤包括：在重复步骤结束时剩余的第 2 种语言的句子之中，具有最大的类似度的句子作为对于输入文的译文输出的步骤。

这样，通过将具有最大类似度的句子作为对于输入文的译文输出，对于输入文来说，作为译文得到最接近于理想的译文的可能性增高。
20

有关本发明第 2 方面的记录介质，记录了让计算机动作的机器翻译计算机程序，在计算机执行该程序时将实施以下机器翻译方法，机器翻译方法采用包含多个由第 1 种语言的句子和第 2 种语言的句子构成对译的对译文集，将第 1 种语言输入文翻译为第 2 种语言，包括以下步骤：摘录步骤，
25 接收第 1 种语言输入文，从对译文集之中，摘录出和输入文之间具有规定关系的、和第 1 种语言的句子成对的第 2 种语言的句子；计算步骤，对于摘录出的第 2 种语言的句子，应用预先决定的多个变形之中的任意变形，计算由变形得到的句子的类似度；选择步骤，从由变形得到的句子中，选

择具有满足规定条件的类似度的句子；重复步骤，直到预先决定的结束条件成立为止，对于选择步骤选择的句子的每一个，重复摘录步骤、计算步骤以及选择步骤；和输出步骤，在重复步骤结束时剩余的第2种语言的句子之中，具有满足预先决定的选择条件的类似度的句子作为对输入文的译文并输出。

通过使该记录介质中存储的程序在计算机中执行，能够在计算机中执行上述机器翻译方法。

通过结合附图进行以下的发明的详细说明，就会清楚本发明的目的、特征、方面以及效果。

10

附图说明

图1表示有关本发明一实施例的机器翻译系统的功能框图。

图2表示图1所示的初始备用选择部32的更详细的功能框图。

图3表示图1所示的备用修正部36的更详细的功能框图。

15

图4表示图3所示的排列搜索部74的处理的详细过程的示意图。

图5表示实现有关本发明一实施例的机器翻译系统的计算机的外观图。

图6表示图5所示的计算机的框图。

图7表示单词排列一例的图。

20

具体实施方式

作为并非象束搜索（beam search）那样对每个单词都进行翻译的系统的机器翻译系统，存在基于范例的系统（范例翻译）。范例翻译是基于对译文集的翻译方式的一种。在对译文集之中，保存了多个由第1种语言的句子和第2种语言译文构成的对译文。如果输入了第1种语言的输入文，在对译文集之中找出和输入文类似的第1种语言的句子，基于找出的第1种语言的句子的译文（第2种语言的句子）生成输出文。

本实施例的机器翻译系统是基于组合了此范例翻译系统和统计机器翻译系统的新框架的系统。

（构成）

30

图 1 表示有关本实施例的机器翻译系统 20 的框图。参照图 1，此机器翻译系统 20 包括：包含多个由第 1 种语言（将此作为语言 J）句子和第 2 种语言（将此作为语言 E）句子构成的对译文的对译文集 34；接收第 1 种语言的输入文 30，为了从对译文集 34 中选择和输入文 30 相似的规定数量（例如 5 个）的第 1 种语言的句子的初始备用选择部 32。

机器翻译系统 20 进一步包括：在统计机器翻译中通常使用的、第 2 种语言的语言模型（P（E））38 以及翻译模型 40（P（J|E））；对于由初始备用选择部 32 选择的多个第 1 种语言的句子的每一个，如后述那样在搜索的同时进行第 2 种语言译文的修正，将采用语言模型 38 以及翻译模型 40 算出的类似度最高的译文作为对输入文 30 进行翻译的结果得到的输出文 42 输出的备选修正部 36。

图 2 表示初始备用选择部 32 的详细框图。参照图 2，初始备用选择部 32 包括 tf/idf 计算部 50，其参照对译文集 34，计算作为表示输入文 30 和对译文集 34 的第 1 种语言的句子的各个之间的类似程度的类似尺度的 tf/idf 标准 $P_{tf/idf}$ 。tf/idf 标准 $P_{tf/idf}$ 是将对译文集 34 的第 1 种语言各句子作为一个文本，利用在信息检索算法中一般采用的被称为文本频率的概念由下面的公式定义的尺度。

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

式中， J_0 表示输入文， $J_{0,i}$ 表示输入文 J_0 的第 i 个单词， $df(J_{0,i})$ 表示对于输入文 J_0 的第 i 个单词 $J_{0,i}$ 的文本频率， N 表示对译文集之中的全部对译文数目。所谓文本频率 $df(J_{0,i})$ 是指输入文 J_0 中的第 i 个单词 $J_{0,i}$ 出现的文本（本实施例中指句子）的数目。

初始备用选择部 32 进一步包括：编辑距离计算部 52，其对于在对译文集 34 中包含的各对译文（ J_k, E_k ）的第 1 种语言的句子 J_k ，进行和输入文 J_0 之间的 DP（Dynamic Programming：动态编程）匹配，计算编辑距离 $dis(J_k, J_0)$ ；和分数计算部 54，其基于由 tf/idf 计算部 50 算出的 tf/idf 标准 $P_{tf/idf}$ 以及由编辑距离计算部 52 算出的编辑距离，根据后述的公式计算各句子的分数。

由编辑距离计算部 52 算出的编辑距离 $dis(J_k, J_0)$ 由以下的公式表示。

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

式中, k 是 $1 \leq k \leq N$ 的整数, $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 以及 $S(J_k, J_0)$ 分别为从句子 J_0 变换为句子 J_k 为止时的插入/删除/置换的次数。

由分数计算部 54 算出的分数 Score 由以下的公式表示。

$$score = \begin{cases} (1.0 - \alpha)(1.0 - \frac{dis(J_k, J_0)}{|J_0|}) + \alpha P_{f/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ 时}) \\ 1.0 & (\text{其它情况}) \end{cases}$$

式中, α 为调整参数, 在本实施例中 α 为 0.2。

参照图 2, 初始备用选择部 32 进一步包含基于由分数计算部 54 算出的分数 score, 选择分数大的规定个数 (本实施例中为 5 个) 的对译文, 作为备选对译文 58 输出, 提供给图 1 所示的备选修正部 36 的对译文选择部 56。

图 3 表示图 1 所示的备选修正部 36 的详细框图。参照图 3, 备选修正部 36 包含: 接收由初始备用选择部 32 输出的备选对译文 58 中所包含的各初始备选对译文 (J_k, E_k), 对于各个初始备选对译文, 使用语言模型和翻译模型, 计算第 1 种语言输入文和第 2 种语言的句子之间类似度最高的维特比排列的维特比排列部 70。由维特比排列部 70 计算对于输入文 J_0 和各个备选对译文 (J_k, E_k) 之中由第 2 种语言的句子 E_k 构成的新的备选对译文 (J_0, E_k) 的初始排列 A_k , 表示排列结束的备选对译文 (J_0, A_k, E_k)。

备选修正部 36 进一步包含: 一致检测部 72, 由维特比排列部 70 判断排列 A_k 算出的排列完毕的各备选对译文 (J_0, A_k, E_k) 之中是否有分数为 1 的备选对译文, 如果一致就将第 1 值, 如果不一致就将第 2 值分别作为一致检测信号 73 输出, 同时如果存在分数为 1 的备选对译文时, 将此备选对译文 75 排列同时输出; 和排列搜索部 74, 当一致检测部 72 没有检测出一致时, 对于来自维特比排列部 70 的排列完毕的备选对译文 (J_0, A_k, E_k) 进行以下所述的排列 A_k 以及译文 E_k 的修正, 最终作为对于输入文 30 的译文, 将表示最高类似度的对译文 77 和其排列同时输出。排列搜索部 74 在此搜索时, 如后述那样使用语言模型 38 以及翻译模型 40。另外一致检测部 72 当检测出一致时, 由排列搜索部 74 停止排列搜索, 使其不

执行。

备选修正部 36 进一步包含译文选择部 76, 对一致检测部 72 输出的一致信号 73 进行应答, 一致检测信号 73 根据第 1 值或者第 2 值选择一致检测部 72 的输出对译文 75 或者排列搜索部 74 的输出对译文 77 之中的任何
5 一个, 作为输出文 42 输出。

图 4 表示排列搜索部 74 进行的修正备选文的搜索和登山法的概略。参照图 4, 排列搜索部 74 包含操作适用部 81A、81B、..., 对于来自维特比排列部 70 的备选对译文 58 中所包含的排列完毕的对译文 80A~80N, 通过应用表示单词的移动/删除/置换等操作之一, 修正排列, 生成多个新的
10 的备选对译文群 82A、82B、...。排列搜索部 74 进一步包含选择处理部 84A、84B、..., 对于这样得到的备选对译文群 82A、82B、...的各个中包含的排列修正后的对译文的各个, 采用语言模型 38 以及翻译模型 40 算出类似度, 从各备选对译文中类似度最高的对译文开始, 按照顺序保留规定的个数(本实施例中为 5 个)将其它的备选对译文删除, 为了从备选对译文
15 文群 82A、82B、...中生成新的备选对译文群 86A、...、86N 的根据类似度。

排列搜索部 74 的操作适用部 81A、81B、..., 例如对于备选对译文群 86A 中所包含的备选对译文 88A、...、88N 也进行上述的动作生成新的备选对译文 90A、...、90N。然后根据类似度的选择处理部 84A、84B、...对于
20 于这些备选对译文群 94A、...、94N 也采用语言模型 38 以及翻译模型 40, 将具有最高类似度的备选对译文 96A、...、96N 保留下来, 生成新的备选对译文群 94A、...、94N。

这样, 排列搜索部 74 将最初的备选对译文 58 中所包含的备选对译文 80 A、...、80N 作为种子(seed), 将对译文的排列不断地应用操作, 生成新的备选对译文。排列搜索部 74 按照根据类似度的选择部 84A、84B、...
25 进行备选对译文的选择时, 对于备选对译文, 当判断计算的类似度不能得到改善的时刻将上述的重复处理中止(登山法)。

这样, 排列搜索部 74 进行对译文的排列的搜索、修正, 根据登山法将在搜索处理中求出的备选对译文和排列之中类似度最高的翻译文作为
30 输出文 42 输出。

操作适用部 81A、81B、…对于排列完毕的对译 (J_0, A_k, E_k) 进行的操作和 Germann 等所记载的几乎相同, 如以下。

(1) 单词的翻译

对单词 J_{0j} 将排列的输出的单词 E_{A_j} 变换为单词 e 。如果单词 e 是
5 NULL, 单词 J_{0j} 被配列为 NULL, $A_j=0$ 。如果单词 E_{A_j} 的派生为 0, 此单词 E_{A_j} 被删除。单词 e 通过反向利用词汇模型计算, 从单词备选中选择。

(2) 单词的翻译以及插入

进行单词的翻译, 在适当的位置插入派生为 0 的一连串的单词。派生
为 0 的一连串的单词的备选, 是对于学习文集从维特比排列中选择的。

10 (3) 单词的翻译以及排列

排列中的第 j 个单词 E_j 移动到第 i 个, 将第 i 个单词 E_i 修正为单词 e 。

(4) 排列的移动

此操作不是进行输出单词顺序的变更, 而是通过排列 A 的移动以及交
换只对排列进行修正。

15 (5) 段的交换

译文 E 之中, 没有重复的部分相互交换。即将从第 i_0 个单词开始到第
 i_1 个单词组成的段和从第 i_2 个单词开始到第 i_3 个 (其中, $i_1 < i_2$) 单词组成
的段进行交换。

(6) 单词 (群) 的删除

20 从译文 E 中将派生为 0 的单词序列删除。

(7) 单词的组合

当译文 E_i 以及 E_i' 中所包含的单词的派生任何一个都为 1 以上时, 将
此两个单词组合。

此 7 种操作之中, 除去 (3)、(4) 剩下的 5 种和由 Germann 等提出
25 的几乎相同。(3) 和 (4) 的操作是在本实施例中新追加的。最初由维特
比排列部 70 执行的维特比排列中, 将第 1 种语言的句子之中没有和第 2
种语言的句子相对应的译文的单词和 NULL 相对应, 或者通过提高派生将
其排列到不适当的单词中。通过操作 (3) 单词的翻译以及排列, 采用词
汇模型能找出每个适当的单词的翻译, 此排列是强制地移动到其它的单词
30 处。同样地, 操作 (4) 排列的移动是通过将已经存在的排列进行移动得

到同样的效果。

(动作)

5 机器翻译系统 20 是采用以下这样的动作。在对译文集 34 中预先包含由第 1 种语言的句子和第 2 种语言的译文构成的多个对译文。另外语言模型 38 以及翻译模型 40 通过某些方法预先准备。

参照图 1，输入文 30 输入到初始备用选择部 32 中。参照图 2，初始备用选择部 32 的 tf/idf 计算部 50 计算输入文 30 和对译文集 34 之中全部的对译文中的第 1 种语言的句子之间的 tf/idf 标准 $P_{tf/idf}$ 。编辑距离计算部 52 同样地计算输入文 30 和全部的对译文集 34 中的第 1 种语言的句子 J_k 10 之间的编辑距离 $dis(J_k, J_0)$ 。

分数计算部 54 利用由 tf/idf 计算部 50 算出的 tf/idf 标准 $P_{tf/idf}$ 以及由编辑距离计算部 52 算出的编辑距离 $dis(J_k, J_0)$ ，根据下式计算上述的分数 score。

$$score = \begin{cases} (1.0 - \alpha)(1.0 - \frac{dis(J_k, J_0)}{|J_0|}) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ 时}) \\ 1.0 & (\text{其它情况}) \end{cases}$$

15 对译文选择部 56 选择对译文集 34 中所包含的对译文之中分数 score 高的规定个数的对译文，作为备选对译文 58 提供给图 3 的维特比排列部 70。

20 参照图 3，维特比排列部 70 将作为输入文 J_0 提供的备选对译文 58 中所包含的对译文 (J_k, E_k) 的各个之中的第 2 种语言的句子 E_k 计算维特比排列 A_k ，将结果以 (J_0, A_k, E_k) 的形式提供给一致检测部 72 以及排列搜索检测部 74。

一致检测部 7 判断来自维特比排列部 70 的对译文之中是否有分数 $score=1$ 的对译文。即一致检测部 72 判断备选对译文之中第 1 语言的句子是否有和输入文一致的，当存在一致的对译文时，一致检测部 72 将一致 25 检测信号 73 的值作为第 1 值，此外情况下，一致检测部 72 将一致检测信号 73 的值作为第 2 值。当存在一致对译文的情况下，一致检测部 72 还将此对译文作为对译文 75 提供给译文选择部 76。

排列搜索部 74 将由维特比排列部 70 提供的排列完毕的备选对译文

(J_0 , A_k , E_k) 作为最初的种子, 参照语言模型 38 以及翻译模型 40 进行上述搜索, 根据登山法进行直到得到类似度最高的译文为止的搜索。搜索的过程中, 排列搜索部 74 对于全部的排列完毕的备选对译文, 采用可能的全部的操作生成新的备选对译文 (和排列)。排列搜索部 74 进一步只留下这样生成的备选对译文 (和排列) 之中满足规定条件的备选对译文 (从类似度最高的开始, 规定个数的对译文), 删除其它的对译文。排列搜索部 74 进一步将剩余的备选对译文作为种子重复同样的处理。然后, 对于生成的备选对译文直到计算的类似度没有改善为止, 结束其路径的搜索 (登山法)。

10 这样, 在对于全部的路径的搜索结束的时刻, 将具有最高类似度的对译文作为最终的输出。排列搜索部 74 将此对译文 77 提供给译文选择部 76。译文选择部 76 当一致检测部 73 为第 1 值时, 将作为一致检测部 72 的输出的对译文 75, 其它的情况下将作为排列搜索部 74 的输出的对译文 77, 分别选择并作为输出文 42 输出。

15 (评价)

进行了对关于上述的实施例的系统的翻译精度的评价。作为文集, 采用了申请人准备的旅行会话用的文集。此文集包含日语、英语、韩国语以及中文的对译文。文集的统计信息在下面的表中表示。

表 1

	中文	英语	日语	韩国语
句子数	167,163			
词语数	956,732	980,790	1,148,428	1,269,888
词汇大小	16,411	15,641	21,896	13,395
单音数	5,207	5,547	9,220	4,191
3 元语法复杂度	45.33	35.35	24.06	20.34

20 将文集全体分为 3 部分。即将 152,169 的句子用于翻译模型以及语言模型的学习中, 将 4,849 的句子用于文本, 剩余的 10,148 用于参数调整。作为参数有学习时的重复的结束标准以及解码的参数调整等的参数。

作为语言模型进行了 4 种语言的 3 元语法 (Tri-gram) 语言模型的学习, 如上面的表中所示, 由复杂度 (perplexity) 尺度进行评价。对于 4 种语

言的组合的全部，关于双向翻译进行了全部的12种的翻译模型的学习。

下表中表示采用有关本实施例的系统在上述的4种语言间进行翻译得到的结果。表中所使用的省略语的意思如下。

5 <WER>表示 Word-error-rate (单词错误率)。这是反映了与标准的模范翻译相比较时的编辑距离(插入/删除/置换)。

<PER>表示位置独立(Position-independent)的WER。这是除了位置的问题，只反映了关于插入和删除的编辑距离。

10 <BLEU>表示 BLUE 分数。这是表示翻译结果的 n-克之中，在成为标准的模范翻译中找出的比率。上述的 WER 以及 PER 不同，BLUE 分数越高翻译的质量越高。

15 <SE>表示主观的评价。分为 A~D (A: 非常完美、B: 基本没有问题、C: 还可以、D: 不合理)的4个等级，由说此母语的人进行的评价。此评价由说母语的人评价为A的比例(A)、评价为A或者B的比例(A+B)、评价为A、B或者C的比例(A+B+C)表示。在本实验中，上述4种语言之中，对于英语之外的3种语言翻译为英语，以及日语以外的3种语言翻译为日语，作为日语-英语之间的翻译结果进行了评价。在此表中，细体表示根据采用束搜索的机器翻译装置的翻译结果，粗体表示根据有关本实施例的机器翻译装置的翻译结果。

表 2

	WER%		PER%		BLEU%		主観的評価(SE)%					
							A		A+B		A+B+C	
C-E	45.0	34.3	39.8	30.3	43.6	56.7	48.4	65.0	65.9	76.9	71.4	81.0
C-J	35.7	25.5	31.3	22.6	56.9	67.8	50.8	69.0	59.4	74.3	66.9	80.2
C-K	38.4	29.1	34.2	26.2	56.1	65.0	-	-	-	-	-	-
E-C	45.0	38.0	39.7	33.4	42.1	51.9	-	-	-	-	-	-
E-J	34.2	29.0	30.5	26.1	59.2	65.7	55.8	65.1	62.4	71.6	70.2	77.8
E-K	38.7	35.6	34.3	31.6	57.3	61.5	-	-	-	-	-	-
J-C	46.8	33.0	38.9	27.8	39.7	57.1	-	-	-	-	-	-
J-E	42.9	35.0	37.4	30.3	47.6	57.4	50.8	63.7	65.7	74.5	70.2	77.6
J-K	27.7	20.8	25.4	19.2	67.2	73.5	-	-	-	-	-	-
K-C	41.9	32.9	34.4	27.6	45.1	55.5	-	-	-	-	-	-
K-E	45.1	36.4	38.5	32.1	44.3	56.8	49.2	61.6	65.7	72.9	72.2	78.4
K-J	26.8	20.8	24.6	19.3	64.3	70.8	56.5	69.2	66.5	77.5	78.4	84.7

20 从此表中能够明白，根据本实施例的机器翻译装置的翻译结果，对于全部的语言的组合，对于翻译为任何一种语言都得到比由束搜索方式的翻

译装置更好的结果。此差异是非常大的，根据本实施例的机器翻译装置的性能和采用束搜索的装置相比是非常高的，这点已经明确了。另外根据本实施例的翻译结果是安定的，限于局部的最适解的情况很少。其原因被认为作为最初的解，检索出和输入文最接近的，以此作为出发点通过进行和
5 上述的登山法组合的搜索，在较大的范围内找出接近于最合适的翻译结果的可能性高。

根据初始备用选择部 32 的初始备用的选择中，如果能够根据某种标准找出和输入文 30 接近的第 1 种语言的句子的对译文，也可以采用和本实施例中采用的初始备用选择部 32 不同的构成的装置。另外，也存在在
10 对译文集 34 中找不到以句子为单位的输入文相对应的译文的情况。在这样的以句子为单位的检索不能实现的情况下，采用比输入句子更小的单位，例如分割为词组或者节等，以分割后的单位通过检索对译文集 34 找出和其相对应的译文，通过组合译文生成初始备选也可以。

进一步，代替由初始备用选择部 32 的初始备用的选择，采用根据其
15 它的某一种翻译方式的翻译装置翻译输入文，将其作为初始备选也可以。例如，作为这样的翻译装置采用范例翻译装置，将范例翻译的结果作为初始备选也可以。此时，范例翻译中所使用的文集可以是和对译文集 34 同样的，也可以是完全不同的文集。

另外，在上述实施例中，排列搜索部 74 的登山法中，采用了和幅度
20 优先搜索同样的算法。但是本发明并非仅限于这样的实施例，采用深度优先的算法在理论上也是可行的。

（由计算机实现）

另外有关本实施例的机器翻译装置也能够通过计算机硬件、在该计算机硬件上动作的程序以及该计算机的存储装置中存储的对译文集、翻译模
25 型以及语言模型实现。特别是，图 4 所示的由排列搜索部 74 进行的搜索通过采用循环编程能够更有效地执行。

根据上述的实施例的说明，这样的程序如果是该行业的一般技术人员，应该能够更容易地实现。

图 5 表示实现此机器翻译装置的计算机系统 330 的外观。图 6 是表示
30 计算机系统 330 的内部构成。

参照图 5，此计算机系统 330 包括：具有 FD（软盘）驱动器 352 以及 CD-ROM（光盘读取专用存储器）驱动器 350 的计算机 340、键盘 346、鼠标 348 和监视器 342。

参照图 6，计算机 340 除了 FD 驱动器 352 以及 CD-ROM 驱动器 350 之外，还包括：CPU（中央处理器）356；连接 CPU356、FD 驱动器 352 以及 CD-ROM 驱动器 350 的总线 366；存储引导程序等的读取专用存储器（ROM）358；与总线 366 相连、存储程序命令、系统程序以及计算数据等的随机存取存储器（RAM）360。计算机系统 330 进一步包括打印机 344。

10 在此处虽然未画出，计算机 340 也可以进一步包括连接局域网（LAN）的网卡。

计算机系统 330 中，为了执行机器翻译装置的动作的计算机程序是存储在插入 CD-ROM 驱动器 350 或者 FD 驱动器 352 中的 CD-ROM362 或者 FD364 中，进一步传送到硬盘 354。另外，程序通过图中未画出的网络向计算机 340 发送，存储在硬盘 354 中也可以。程序执行时下载到 RAM360 中。也可以从 CD-ROM362、FD364 或者通过网络直接将程序下载到 RAM360 中。

此程序包含计算机 340 执行作为此实施例的机器翻译装置的动作的多个命令。执行此方法所必要的几个基本功能由在计算机上动作的操作系统（OS）或者第三方的程序、或者由计算机 340 中所安装的各种工具包的模块提供。也就是说，此程序并非必须包括实现此实施例的系统以及方法所必要的全部功能。此程序的命令之中，只需包括为了能得到所期望的结果，以控制的方法通过读出适当的功能或者“工具”，实现上述的机器翻译装置

20 的命令即可。因为计算机系统 330 的动作是周知的，此处不重复叙述。

此次展示的实施例只是示例，本发明并非仅限于上述实施例。本发明的范围是在参考发明的详细说明书的记述的基础上，包括由权利要求书的各个权利要求所提到的、和其中所述具有相同的意思以及范围内全部的变更。

25

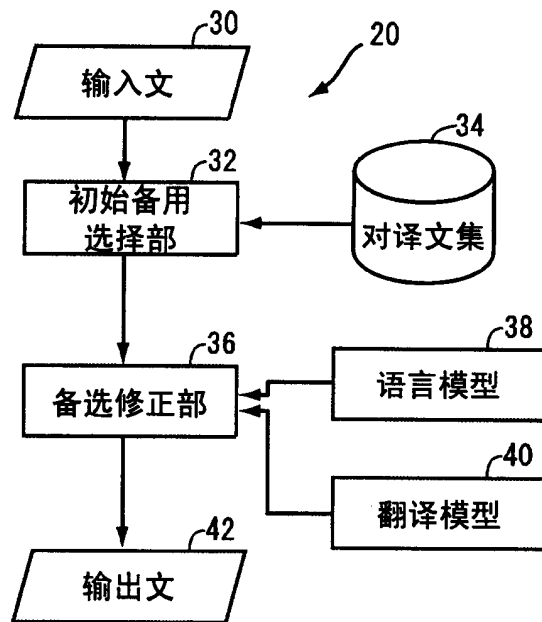


图 1

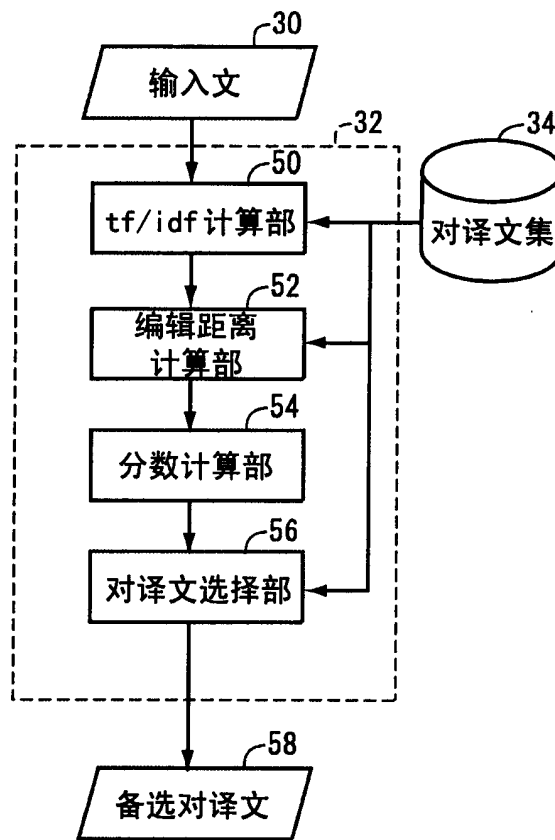


图 2

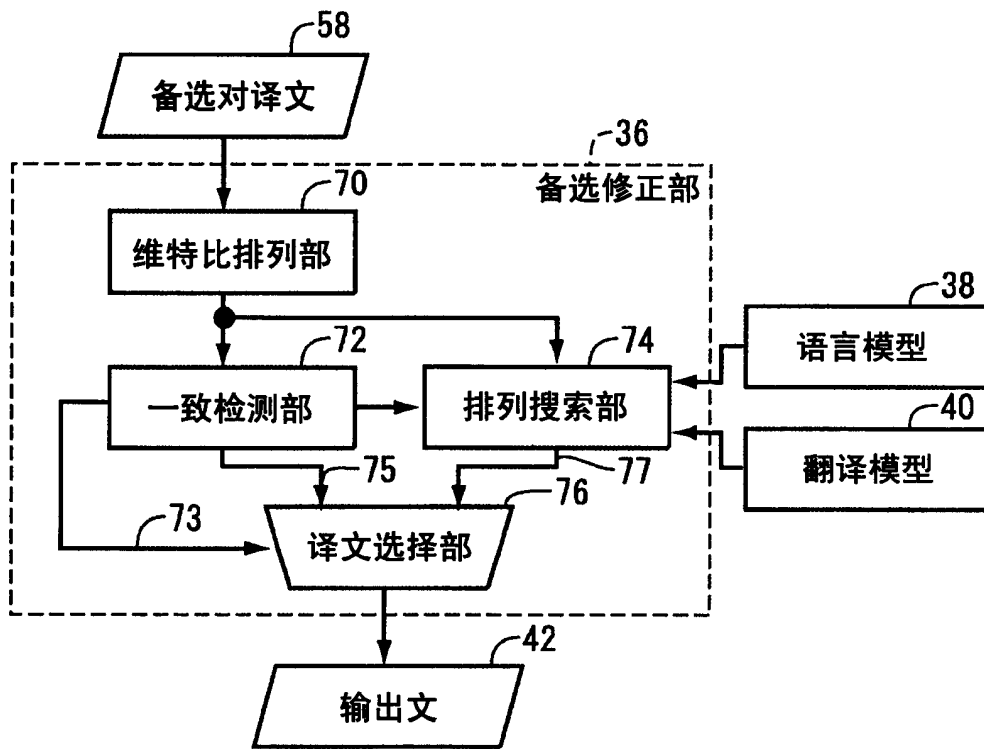


图 3

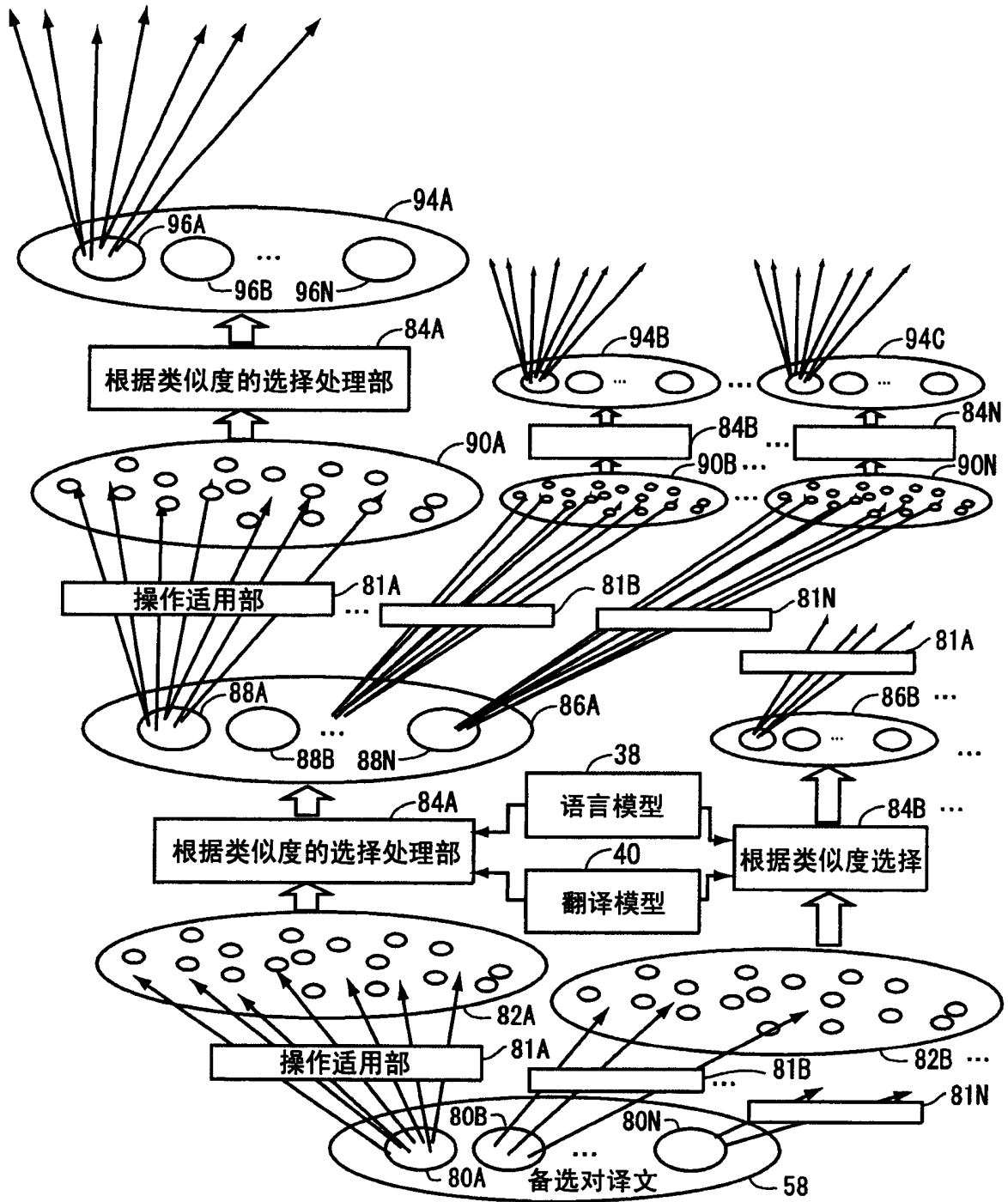


图 4

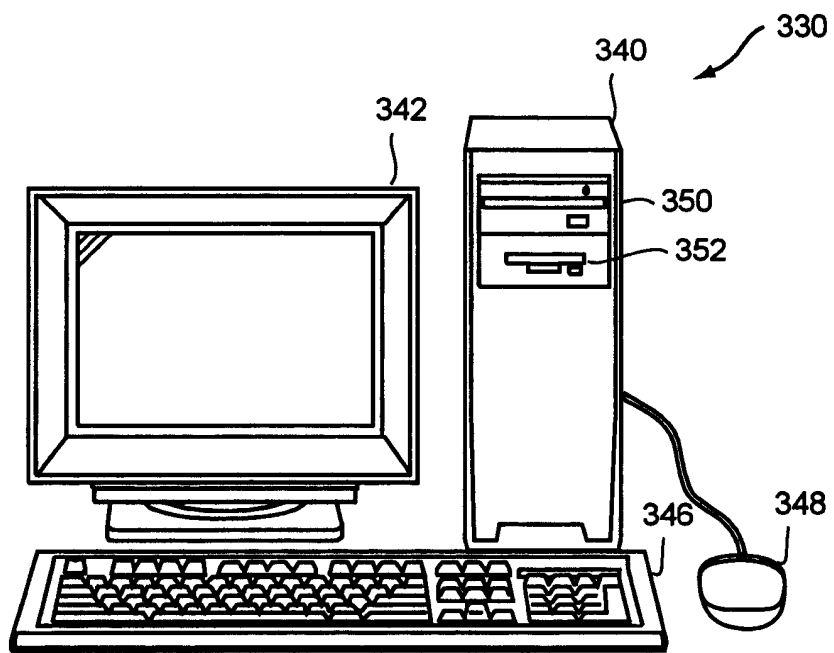


图 5

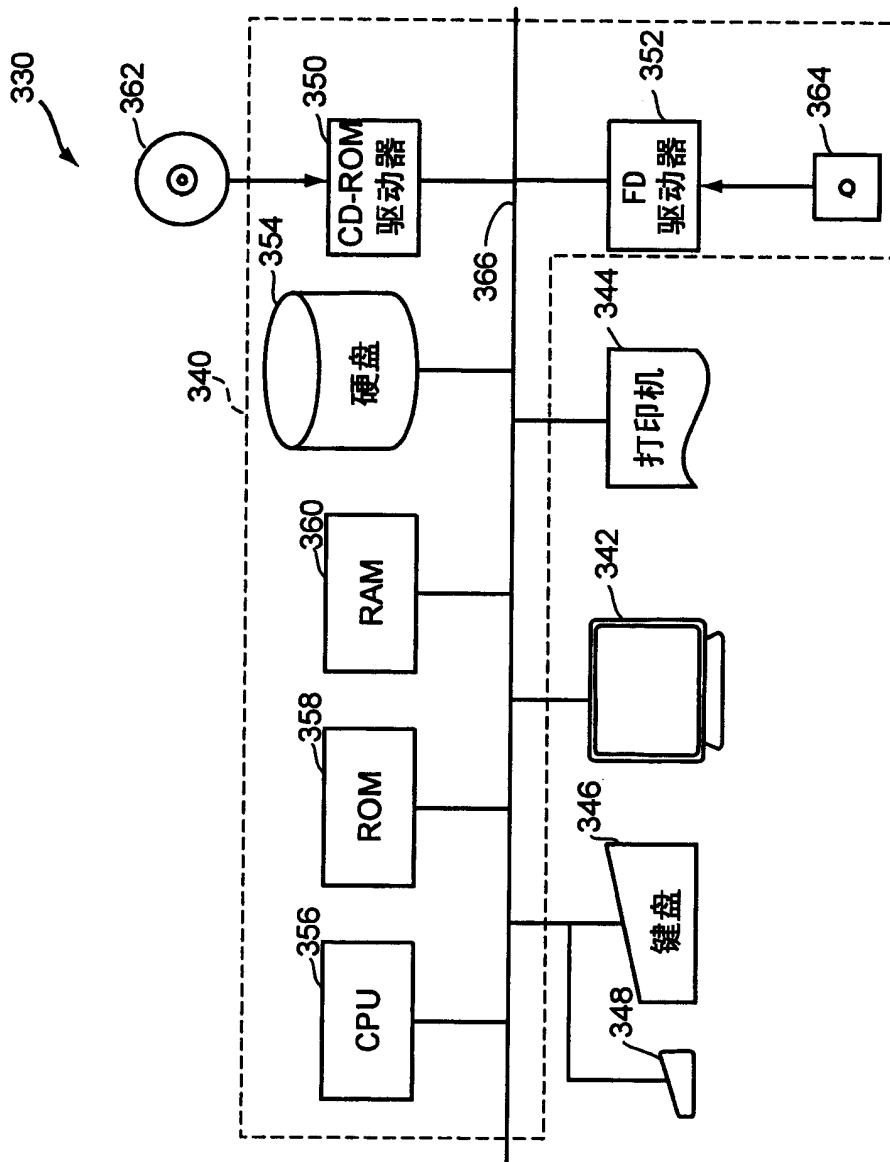


图 6

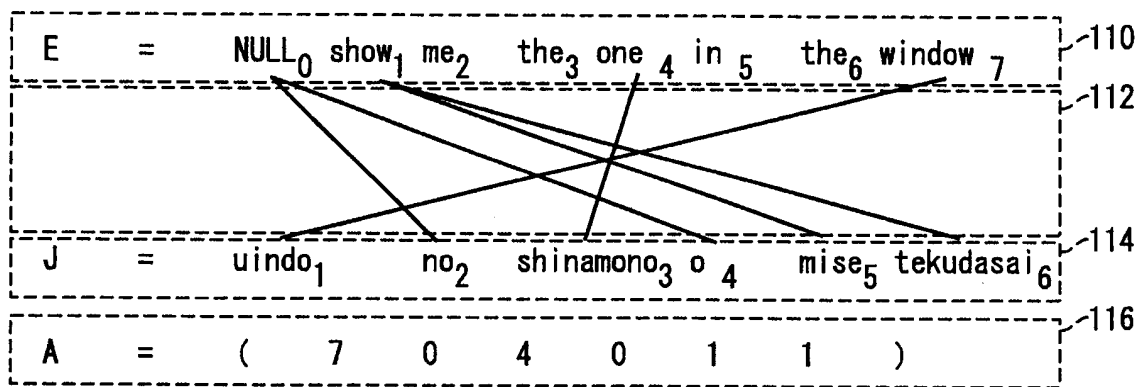


图 7